DEVELOPMENT OF A NOVEL METHOD FOR GENOME-WIDE IDENTIFICATION OF PROTEINS EXPORTED DURING INFECTION AND FUNCTIONAL STUDIES OF ONE *IN VIVO* EXPORTED PROTEIN IN *MYCOBACTERIUM TUBERCULOSIS*

Ellen Foot Perkowski

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Microbiology and Immunology in the School of Medicine.

Chapel Hill
2015

Approved by:

Miriam Braunstein

Thomas Kawula

Bill Goldman

Peggy Cotter

Anthony Richardson

ABSTRACT

Ellen Foot Perkowski: Development of a Novel Method for Genome-wide Identification of Proteins Exported during Infection and Functional Studies of One *in vivo* Exported Protein in *Mycobacterium tuberculosis*
(Under the direction of Miriam Braunstein)

Intracellular pathogens manipulate and outwit the host's immune defenses and reprogram the hostile intracellular environment into a hospitable replicative niche. The intracellular pathogen *Mycobacterium tuberculosis* is responsible for the disease tuberculosis, which kills approximately 1.5 million people per year. *M. tuberculosis* produces many proteins that are exported: transported out of the bacterial cytoplasm to the bacterial cell surface and out into the host environment. Exported proteins are located at the host-pathogen interface, in an ideal location to manipulate the host response and allow for intracellular growth, and exported proteins contribute significantly to virulence. Unfortunately, approaches used to identify proteins exported by the bacteria are limited to bacteria growing in laboratory media (*in vitro*). Because *in vitro* conditions cannot mimic the complexity of the host environment, there are likely critical exported virulence factors that have been missed because they are only exported in the context of infection. The main objective of the research described in this dissertation was to develop a method to identify proteins that are exported by bacterial pathogens during infection of a host (*in vivo*). We developed a novel method that we refer to as EXIT, EXported In vivo Technology, and applied it to identify *M. tuberculosis* proteins exported during murine infection. EXIT identified 593 *in vivo* exported proteins, 100 that were experimentally shown to be exported for the first time, and 32 proteins with no *in silico* predicted export signals. EXIT identified 38

proteins exported significantly more *in vivo* than *in vitro*, suggesting that temporal or spatial control of their export is important to infection. 21 of these 38 proteins have unknown function, making them particularly interesting for future functional characterization. We focused on one of the EXIT identified *in vivo* exported proteins, OmasA, a protein of unknown function for further study. We demonstrated that OmasA was required for *M. tuberculosis* virulence in a mouse model of tuberculosis. We further demonstrated a function for OmasA in stabilizing multi-protein Mce transporters required for lipid import. Future studies will focus on assigning function to new EXIT identified exported proteins, in particular proteins whose export is significantly induced during infection.

May our next journey be as wonderful as the last.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

Φ	hydrophobic amino acid
°C	degrees Celsius
µL	microliter
ABC	ATP-binding cassette
ADS	Albumin Dextrose Saline
Ag	antigen
ANL	azidonorleucine
ANOVA	analysis of variance
ATP	adenosine triphosphate
BCG	bacillus Calmette-Guerin attenuated vaccine strain
Bla	β-lactamase
Bp	nucleotide base pairs
BSL-3	Biosafety level 3
C-	carboxy terminus of a protein
C14	carbon-14
CD	cluster of differentiation
cDNA	complementary DNA
CFP	culture filtrate protein
cfu	colony forming units
Cl	chlorine
Cs	cesium
CW	cell wall

D	aspartic acid
dATP	deoxyadenosine triphosphate
dH_2O	distilled deionized water
DNA	deoxyribonucleic acid
Dpi	days post infection
E	glutamic acid
Ecc	ESX conserved component
ESAT-6	early secreted antigenic target of 6 kDa
Esp	ESX specific protein
ESX	ESAT-6 secretion system
EXIT	EXported In vivo Technology
FDR	false discovery rate
g	grams
g	gravity
GC	guanine and cytosine
Glu	glutamic acid
GR	genetic reporter
H	hydrogen
HA	Influenza hemagglutinin epitope
HAI	histological activity index
HIV	Human Immunodeficiency Virus
HRP	horseradish peroxidase
Hsp	heat shock protein

Hyg	hygromycin
ICM/Dot	intracellular multiplication/defective organelle trafficking
IVET	*in vivo* expression technology
Kan	kanamycin
kb	kilobase
kD	kiloDalton
kDa	kiloDalton
Lab	laboratory
LB	Luria Bertani media
Log10	logarithm base 10
Log2	logarithm base 2
M. smegmatis	*Mycobacterium smegmatis*
M. tuberculosis	*Mycobacterium tuberculosis*
Mas	Mce associated protein
Mce	mammalian cell entry
MEM	membrane
Mg	magnesium
mg	milligram
mL	milliliter
MmpL	mycobacterial membrane protein large
MS	mass spectrometry
Msmeg	*Mycobacterium smegmatis*
Mtb	*Mycobacterium tuberculosis*

N-	amino-terminus of a protein
Na	sodium
NEB	New England Biolabs
NH_4	ammonium
NIH	National Institutes of Health
OD	optical density
omasA	orphaned Mce associated protein A
ORF	open reading frame
p	plasmid
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDIM	phthiocerol dimycocerosate
PE	proline glutamic acid repeat containing protein
PGRS	polymorphic guanine-cytosine rich sequence
PhoA	alkaline phosphatase
P_i	inorganic phosphate
PO_4	phosphate
PPE	proline proline glutamic acid repeat containing protein
Pro	proline
qRT-PCR	quantitative reverse-transcriptase PCR
R	arginine
^{R}	resistant
Ra	*M. tuberculosis* strain H37Rv attenuated

Res	resolvase cassette
RNA	ribonucleic acid
RND	Resistance Nodulation Division
Rv	*M. tuberculosis* strain H37Rv attenuated
S	sensitive
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sec	secretory pathway
SL-1	sulfolipid-1
SOL	soluble
SP	signal peptide
SPI	*Salmonella* pathogenicity island
SRP	signal recognition particle
T3SS	Type III secretion system
Tad	Tight adherence
Tat	Twin-arginine translocation
TM	transmembrane domain
Tn	transposon
Tn-seq	Transposon deep sequencing
TraSH	Transposon site hybridization
tRNA	transfer RNA
Tw	polyoxyethylene sorbitan monoleate
Tween	polyoxyethylene sorbitan monoleate
Ty	Tyloxapol

WCL	whole cell lysate
WT	wild type
X	variable amino acid
Y	tyrosine
Δ	deletion

CHAPTER 1: INTRODUCTION

In 1882 Robert Koch described the bacterium *Mycobacterium tuberculosis* as the causative agent of the disease then known as "consumption" and now better known as tuberculosis (Sakula, 1983). *M. tuberculosis* fulfilled "Koch's postulates", cultures of *M. tuberculosis* could be isolated from patients, purified, and used to infect laboratory animals (Koch, 1932). These animals would then develop characteristic tuberculosis disease, and *M. tuberculosis* bacilli could be isolated from their lung lesions (Koch, 1932). Prior to Koch's discovery, Jean-Antoine Villimen had demonstrated that a rabbit could develop tuberculosis after inoculation of material from a human granuloma; however, many still attributed the disease to poor environmental conditions or genetics (Daniel, 2006). Koch's discovery led to a more broad understanding of the infectious nature of tuberculosis (Daniel, 2006). One of the key symptoms of tuberculosis, prolonged coughing, is now know to be instrumental in allowing the bacteria to escape the lungs of an infected individual and be transmitted through the air to surrounding people. This transmission is so effective that one-third of the world's population (two billion people), are thought to be infected with *M. tuberculosis* (World Health Organization, 2014).

The development of the antibiotic streptomycin in the 1940s led to the first treatment for tuberculosis (Schatz *et al.*, 1944). Even early on, it was clear that tuberculosis was harder to treat than many other bacterial pathogens. Treatment regimens had to be very long or else patients would quickly relapse, often with drug resistant strains (Sakula, 1983). Despite 70 years since the first anti-mycobacterial drug was introduced, we are still struggling with effective treatments for tuberculosis. The World Health Organization estimates that, of the nine million individuals

presenting active tuberculosis each year, approximately 480,000 patients are infected with multi-drug resistant strains (World Health Organization, 2014). In the absence of effective antibiotic treatment tuberculosis is usually fatal (Tiemersma *et al.*, 2011). Today, despite the advances of modern medicine, 1.5 million people continue to die from tuberculosis every year (World Health Organization, 2014).

Upon inhaling *M. tuberculosis*, the bacteria are transferred down into the lungs and into the alveolar spaces, where they can be engulfed by resident alveolar macrophages (Russell *et al.*, 2010). These macrophages are normally the immune system's first line of defense against inhaled material, and will usually engulf and destroy invading bacteria. *M. tuberculosis*, however, has developed many strategies to avoid destruction, and is able to grow inside macrophages, in a compartment called the phagosome (Russell *et al.*, 2010). *M. tuberculosis* reconstructs the phagosome into a hospitable environment, blocking phagosomal maturation and preventing fusion with lysosomes, which would normally lead to destruction of the phagosomal contents (Rohde *et al.*, 2007; Armstrong & Hart, 1971). *M. tuberculosis* will undergo many rounds of replication inside macrophages. Although these macrophages fail to eliminate the *M. tuberculosis* bacilli, they will produce chemoattractant molecules resulting in an influx of inflammatory cells including macrophages, dendritic cells, and neutrophils (Russell *et al.*, 2010; Orme & Basaraba, 2014). Macrophages present antigens to dendritic cells at the site of infection, and these dendritic cells migrate to lymph nodes where they prime T-cells. Antigen specific T cells are then recruited to the site of lung infection where they recognize and activate *M. tuberculosis* infected macrophages, which serves to limit further intracellular bacterial replication, and they organize the inflammatory cells (Russell *et al.*, 2010; Orme & Basaraba, 2014). Eventually, formation of a fibrous and calcified barrier will wall off the inflammatory

focus (Russell *et al.*, 2010; Orme & Basaraba, 2014). This process ultimately leads to formation of a highly specialized structure called a granuloma (Figure 1.1).

Once granulomas have formed, most infected individuals will have no symptoms, a condition known as latent tuberculosis. Although the immune system contains *M. tuberculosis* within granulomas, the bacteria are not fully eradicated and for approximately 10% of infected individuals this containment will ultimately fail during their lifetime (Flynn & Chan, 2001). Individuals who fail to contain the *M. tuberculosis* bacilli will develop active tuberculosis. Although certain risk factors are known, such as immune dysfunction, the cause of this reactivation and the factors determining when it will occur are still mysterious (Flynn & Chan, 2001). The underlying ability for *M. tuberculosis* to cause disease, survive in the macrophage environment and avoid complete destruction by the immune system remains poorly understood.

M. tuberculosis is challenging to treat with antibiotics for several reasons. Drugs must penetrate granulomas and host cells to reach intracellular *M. tuberculosis* bacilli. *M. tuberculosis* bacilli can persist in a non-replicating state, which is resistant to most antibiotics. Additionally, the *M. tuberculosis* cellular structure is unique, and poses further complications to drug treatment. *M. tuberculosis* has both a cytoplasmic membrane and a specialized outer membrane referred to as the myco-membrane (Figure 1.2). The outermost layer of *M. tuberculosis* is a thick capsule surrounding the myco-membrane. The myco-membrane is composed of specialized long chain lipids including mycolic acids. The myco-membrane is covalently linked to arabinogalactan, which is attached to peptidoglycan within the periplasm (Figure 1.2). As well as presenting a challenge to drug entry, the complex cell wall of *M. tuberculosis* would seem to present a challenge to protein exit as occurs during protein secretion.

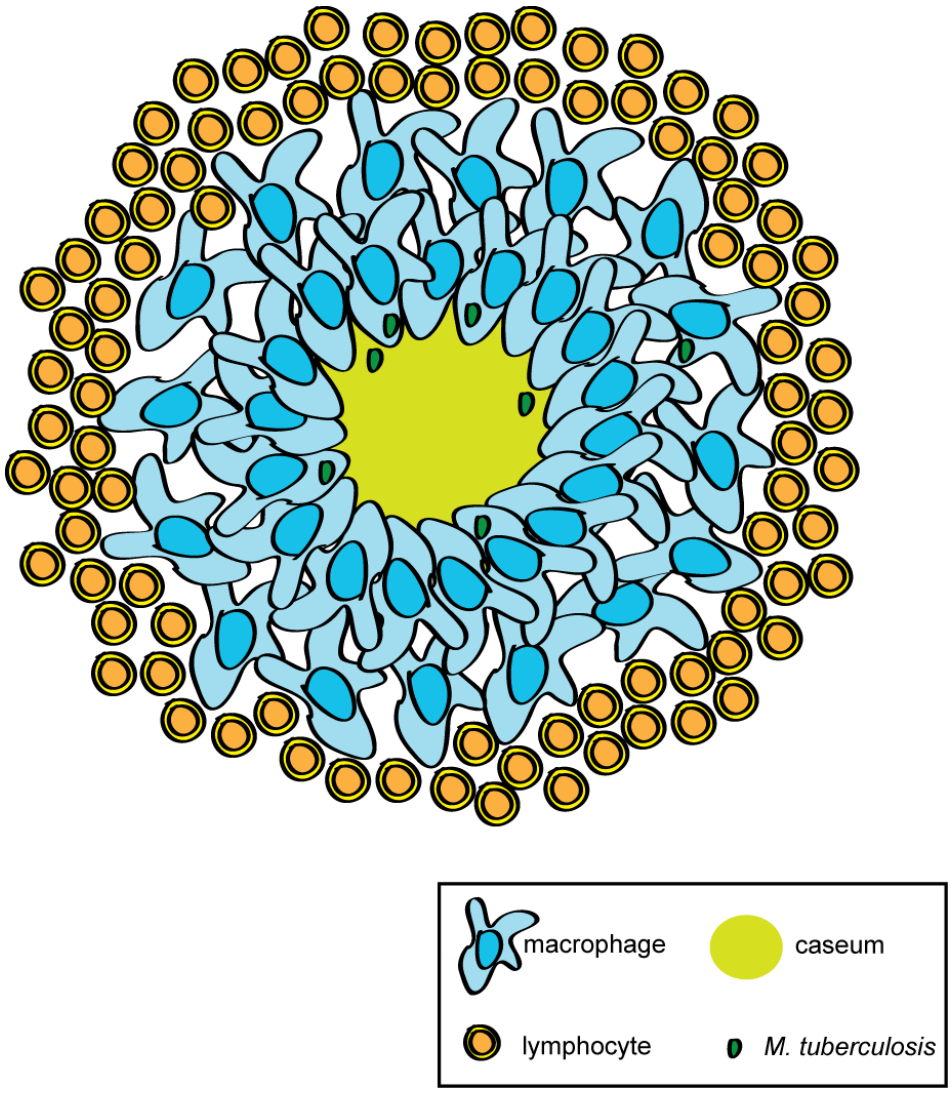

Figure 1.1. Structure of a classic Human Granuloma. Adapted from (Barry *et al.*, 2009). Granulomas consist of a large number of macrophages surrounded by lymphocytes (B cells, $CD4^+$ T cells, $CD8^+$ T cells). Macrophages within granulomas can appear epithelioid or foamy. Mature granulomas in humans become necrotic and hypoxic at the core, which is referred to as the caseum.

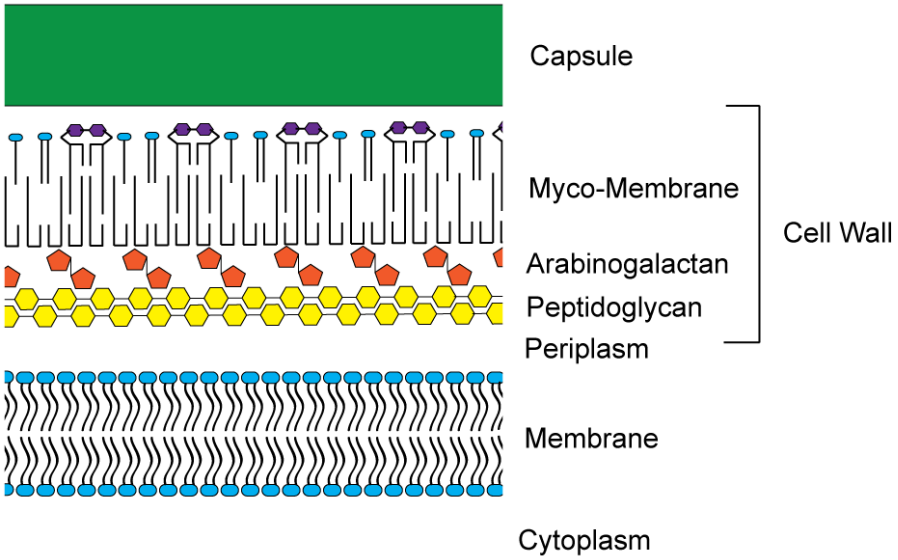

Figure 1.2. Cellular structure of *M. tuberculosis*. *M. tuberculosis* is an acid-fast organism with an cytoplasmic membrane consisting of a phospholipid bilayer (represented in blue) surrounded by a layer of peptidoglycan (yellow). Unique features of the *M. tuberculosis* cell wall include arabinogalactan (orange) which covalently crosslinks the outer myco-membrane to the peptidoglycan. The myco-membrane primarily consists of specialized lipids composed of mycolic acids (Marchand *et al.*, 2012). The mycobacterial capsule is poorly characterized, however, it recently has been visualized by cryo-electron microscopy (Sani *et al.*, 2010).

Bacterial exported proteins are critical to the virulence of intracellular pathogens

We are beginning to understand the important role that exported proteins play in virulence for many bacterial pathogens. This dissertation will refer to exported proteins as any protein produced in the bacterial cytoplasm and actively transported to the bacterial cytoplasmic membrane, cell wall, or out into the host environment (Figure 1.3). Exported proteins lie at the host-pathogen interface, and thus are in an ideal location to interact with the host and contribute to virulence. Many of the best studied virulence factors are exported proteins. For example, virulence of the bacterial pathogen *Legionella pneumophila* requires a type IV secretion system (ICM/Dot), which delivers nearly 300 proteins out of the bacterium and directly into the cytoplasm of host cells (Isaac & Isberg, 2014). Effectors of the ICM/Dot system mediate construction of the *Legionella*-containing vacuole, intracellular replication, and evasion of lysosome fusion (Isaac & Isberg, 2014). *Salmonella enterica*, another well studied bacterial pathogen, relies on exported proteins to attach to host cells, invade, program the fate of its intracellular vacuole, and ultimately replicate to cause infection (Agbor & McCormick, 2011). Given the importance of exported proteins to intracellular pathogens, it is not surprising that *M. tuberculosis* also relies on numerous exported proteins to survive and grow in macrophages and cause disease in a host. The goal of this dissertation was to identify proteins that are exported in the host, and begin the process of assigning function to *in vivo* exported proteins of *M. tuberculosis.*

Mechanisms of protein export in *M. tuberculosis*

Currently there are four types of systems known to export proteins out of the cytoplasm of *M. tuberculosis*: the general Sec pathway, the SecA2-dependent pathway, the Tat (Twin-Arginine Translocation) pathway, and the ESX (Type VII) pathways (Ligon *et al.*, 2012). The

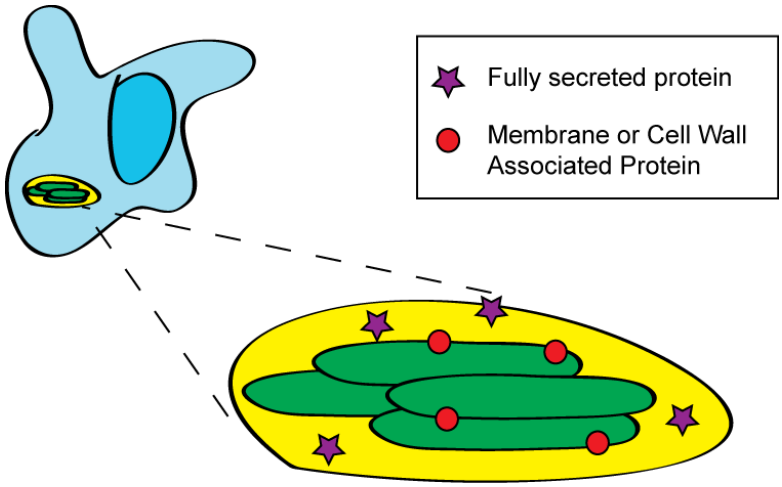

Figure 1.3. Exported Proteins of *M. tuberculosis*. *M. tuberculosis* bacilli (represented in green) grow within the phagosome (yellow) of macrophages (blue). Exported proteins are located at the bacterial-host interface and are in an ideal position to be virulence factors. Exported proteins include those which are fully secreted into the host environment (stars) as well as membrane or cell wall associated proteins.

transport mechanism as currently understood for each of these export pathways, and a review of the general importance of each export pathway to *M. tuberculosis* is described below. Because each of these systems exports multiple proteins, when mutants defective in any one of these pathways are studied the cumulative effect of many exported proteins not being properly localized is observed. Much of what we currently know about the role of *M. tuberculosis* exported proteins in virulence comes not from studying and identifying the role of individual proteins, rather from studying genetic deletions disrupting entire export systems.

The General Secretion System: SecA1 dependent protein export

The general secretion system (Sec) is the housekeeping system that carries out the bulk of export/secretion, as recently reviewed (Ligon *et al.,* 2012). The Sec pathway is essential and is required for the majority of protein export that occurs in all bacteria. Sec exported proteins are synthesized as precursors containing N-terminal signal peptides, which direct them to be post-translationally exported by the Sec pathway (Ligon *et al.*, 2012). SecA1 is an ATPase which recognizes signal peptides of proteins for export, delivers them to the SecYEG membrane channel, and undergoes successive rounds of ATP hydrolysis to transport the protein across the cytoplasmic membrane (Ligon *et al.*, 2012). After export, a signal peptidase cleaves the signal peptide and releases the mature domain of the protein into the periplasm. Sec exported proteins can remain in the cell wall or get fully secreted out of the bacterial cell by a poorly defined second mechanism. As in other bacteria, the Sec system is predicted to export the majority of proteins in *M. tuberculosis*, including proteins required for basic cellular physiology, cell maintenance, and virulence (Ligon *et al.*, 2012). Lipoproteins are a subset of Sec exported proteins that have a lipid moiety attached to their N-terminus responsible for localizing lipoproteins within the cytoplasmic membrane or cell wall. Lipoproteins contain a distinct signal

peptide and are cleaved by a specialized lipoprotein signal peptidase (Ligon *et al.*, 2012). The lipoprotein signal peptidase is required for virulence, highlighting the importance of exported lipoproteins to *M. tuberculosis* (Sander *et al*., 2004). SecA1, SecYE, and the general signal peptidase of mycobacteria are all essential for bacterial cell viability (Ligon *et al.*, 2012).

Co-translational Export of Transmembrane Proteins

The SecYEG channel is additionally important to the export of integral membrane proteins, through a recently reviewed mechanism (Facey & Kuhn, 2010; Ligon *et al.*, 2012). As polypeptide chains emerge from the ribosome they are scanned for hydrophobic domains that may constitute transmembrane domains by the signal recognition particle (SRP). Upon detection of a hydrophobic domain, SRP will bind and transport the growing polypeptide chain and ribosome complex to the SRP receptor FtsY. FtsY will pass the protein to the SecYEG complex, where the protein will undergo co-translational export. Transmembrane domains can pass sideways through a gate in the SecY channel with the help of YidC to stably integrate into the membrane. A subset of transmembrane proteins can be inserted by YidC alone (Facey & Kuhn, 2010). Transmembrane proteins are required for basic cellular physiology, cell maintenance, and virulence; therefore, SRP, FtsY, and YidC are all essential to bacterial cell viability (Ligon *et al.*, 2012).

Accessory Sec Export Pathway: SecA2 dependent protein export

An unusual property of mycobacteria is that they have two non-redundant SecA homologues (SecA1 and SecA2) (Braunstein *et al.*, 2001). The SecA2 protein of mycobacteria is thought to function much like SecA1, binding to SecA2 substrates, transporting them to the SecYEG complex, and transporting preproteins through SecY with rounds of ATP hydrolysis (Ligon *et al.*, 2013; Feltcher & Braunstein, 2012). Some SecA2 substrates have Sec-like signal

peptides, and these signal peptides are required for export (Gibbons *et al.*, 2007). Surprisingly, the signal peptide alone does not confer SecA2 specificity (Feltcher *et al.*, 2013). Undefined elements within the mature domain determine export through the SecA2 pathway, and the leading hypothesis is that the SecA2 pathway in mycobacteria may exist to export proteins that tend to fold prior to export (Feltcher *et al.*, 2013). Furthermore, not all SecA2 substrates have signal peptides, highlighting the importance of additional targeting domains within the mature domain of SecA2 exported proteins (Braunstein *et al.*, 2001; Braunstein *et al.*, 2003).

SecA2 is not essential for *M. tuberculosis* growth *in vitro*, but mutants lacking a functional SecA2 dependent protein export pathway are significantly attenuated (Braunstein *et al.*, 2003; Kurtz *et al.*, 2006). Mice infected with a *ΔsecA2* mutant of *M. tuberculosis* have lower bacterial burden and survive longer than WT *M. tuberculosis* infected mice (Braunstein *et al.*, 2003). Additionally, the *ΔsecA2* mutant is unable to grow in macrophages (Kurtz *et al.*, 2006). These results indicate that one or more proteins exported through the SecA2 dependent protein export pathway are required for virulence. Recent evidence suggests that SecA2 exported proteins play roles in preventing phagosomal maturation (Sullivan *et al.*, 2012) and promoting nutrient acquisition (Feltcher *et al.*, 2015), although the contribution of individual SecA2 exported proteins to attenuation of the *ΔsecA2* mutant has yet to be determined.

Export of Pre-folded Substrates: The Twin-Arginine Translocation Pathway

Mycobacteria use the Twin-Arginine Translocation (Tat) pathway to export pre-folded proteins in a Sec-independent pathway, which has been recently reviewed (Ligon *et al.*, 2012). Tat exported proteins contain distinct N-terminal signal peptides which direct them for export through the TatABC complex. The TatBC membrane protein complex binds Tat precursors and determines whether they are properly folded for export. TatA is recruited to the complex and

forms homo-oligomers of varying sizes to form a channel large enough to transport the pre-folded substrate. Transport of pre-proteins through the TatABC complex is energized by the proton motive force. After export, the signal peptides of Tat exported proteins can be cleaved by signal peptidases (Ligon *et al.*, 2012).

The Tat export pathway has been studied in many bacterial pathogens and is frequently required for export of virulence factors and, thus, contributes to pathogenesis (Ligon *et al.*, 2012). The prediction of Tat exported proteins is clouded by the limited agreement between prediction algorithms; however, as many as 95 proteins may be exported by the Tat pathway in *M. tuberculosis* (McDonough *et al.*, 2008). Characterized Tat substrates include proteins important to both drug resistance and virulence. The Tat pathway is not essential for the *in vitro* viability of most bacterial, thus it is surprising that the Tat pathway is essential to *M. tuberculosis* viability *in vitro* (Saint-Joanis *et al.*, 2006).

Type VII secretion: ESX Specialized Secretion Systems

M. tuberculosis encodes five Type VII secretion systems, named ESX-1 to ESX-5, which are named for the first known exported substrate of these systems, ESAT-6. ESX secretion was recently reviewed in (Houben *et al.*, 2014). ESX gene clusters contain genes encoding ESX conserved components (Ecc), ESX specific proteins (Esp), and mycosins (MycP) which form the export machinery. Some genes encoding ESX substrates are also located within these genomic regions, including genes encoding the classical ESAT-6 like proteins and Pro-Glu and Pro-Pro-Glu repeat containing proteins (PE/PPE) (Houben *et al.*, 2014). The export mechanism of Type VII systems is under active investigation. These systems appear to form multi-protein transport complexes with components in the cytoplasm and cytoplasmic membrane (Houben *et al.*, 2014). Most mycobacterial ESX systems contain cytoplasmic proteins that bind to exported substrates

prior to export and are predicted to function as chaperones (Houben *et al.*, 2014). The exported substrates are thought to be targeted, possibly with the assistance of chaperones, to the large ESX membrane complex consisting of conserved components EccBCDE, all of which are required for ESX secretion; however, the individual role of each protein within the complex remains unclear (Houben *et al.*, 2014). EccD is a multi-membrane spanning protein predicted to form a channel for protein transport (Houben *et al.*, 2014). EccC contains three nucleotide binding domains, and is predicted to energize transport of exported substrates through the complex by ATP hydrolysis (Houben *et al.*, 2014). The subtilisin-like protease, MycP, cleaves at least one ESX substrate and may additionally regulate export in a protease-independent process. Because many ESX substrates are fully secreted into the culture filtrate, it has been hypothesized that the ESX secretion system spans both the cytoplasmic membrane and the outer myco-membrane, to direct proteins out of the bacterium in a single step from the cytoplasm to the extracellular environment (Houben *et al.*, 2014). However, so far, no ESX secretion components have been identified within the myco-membrane.

The full repertoire of exported substrates of each ESX system is still undefined. ESX exported proteins do not contain classical signal peptides; rather, a YxxxD/E conserved motif has been shown to be required for export (Daleke *et al.*, 2012). Surprisingly, the classical substrate ESAT-6 does not contain this motif. ESX substrates are often exported as heterodimers, and often it is only one of the proteins in the complex that contains the YxxxD/E motif (Houben *et al.*, 2014). ESX-1 is the best characterized ESX system in *M. tuberculosis*. ESX-1 has been shown to export a small number of proteins, including EsxA (ESAT-6), EsxB (CFP-10), and EspB (Houben *et al.*, 2014). *Mycobacterium marinum*, a pathogenic mycobacterial species naturally infecting fish, has been used as a model system to study ESX secretion. Approximately

7% of the *M. tuberculosis* proteome are Pro-Glu or Pro-Glu Glu repeat containing proteins (PE/PPE proteins), most of which are thought to be surface localized (Goldberg *et al.*, 2014; Banu *et al.*, 2002; Brennan *et al.*, 2001). In *M. marinum*, the ESX-5 system was shown to export a large number of PE/PPE proteins, including those encoded outside of the ESX-5 locus (Abdallah *et al.*, 2009). However, only a subset of PE/PPE proteins in *M. tuberculosis* are exported through ESX-5, although proteins in *M. tuberculosis* containing PE domains do have the YxxxD/E motif (Bottai *et al.*, 2012). Future studies are necessary to characterize exported substrates of the additional ESX systems.

ESX-1 plays a clear role in *M. tuberculosis* virulence, as shown by attenuation of *esx-1* mutants during mouse and macrophage infection (Pym *et al.*, 2002; Lewis *et al.*, 2003). ESX-2 and ESX-4 have not been previously studied. ESX-3 is essential for *M. tuberculosis* growth *in vitro* due to a role in iron acquisition (Siegrist *et al.*, 2009). Additionally, ESX-5 has recently been shown to be important during *M. tuberculosis* infection, as shown by attenuation of *esx-5* mutants during mouse and macrophage infection (Bottai *et al.*, 2012).

Identification of *M. tuberculosis* exported proteins

Comprehensive study of the role of exported proteins in *M. tuberculosis* virulence first requires a fundamental understanding of which proteins are exported. Currently, *in silico* predictions are most frequently used to predict the exported nature of a protein; however, *in silico* predictions require experimental validation. The approaches currently used to assign a protein as being exported by *M. tuberculosis* are reviewed below, and Chapter 2 of this dissertation describes a novel method we developed to experimentally identify proteins that are exported by *M. tuberculosis* in the context of host infection.

Prediction of Exported Proteins by Bioinformatics (in silico)

The use of computational, or *in silico*, predictions to identify exported proteins is by far the easiest and most common approach. Exported proteins comprise approximately 20% of bacterial proteomes, and include proteins exported by a large diversity of systems. There are a wide variety of programs designed to predict the cellular location of proteins through identifying signal peptides and transmembrane domains, common features of exported proteins. One consistent difficulty with using these prediction programs is that the prediction algorithms have been trained on datasets from Gram positive and/or Gram negative bacteria; none have been trained on GC rich acid-fast mycobacteria. Despite the fact that no protein export prediction tools are optimized for mycobacteria, *in silico* analyses remain the most common method used to predict the subcellular location of a protein in *M. tuberculosis*. However, prediction algorithms are far from perfect, and *in silico* predictions must be followed up with experimentally validation.

Prediction programs for Sec exported proteins inspect the N-terminus of a protein for a signal peptide. Signatures of a Sec signal peptide include a positively charged N-terminus, a hydrophobic core, and a polar region C-terminal domain containing the cleavage site (Feltcher *et al.*, 2013). Programs designed to predict Sec signal peptides include SignalP (Petersen *et al.*, 2011), Psort (Nakai & Horton, 1999), and PrediSi (Hiller *et al.*, 2004). Recent data suggests that Signal P is the most accurate of the available algorithms when predicting signal peptides in mycobacteria (Leversen *et al.*, 2009). Lipoproteins, a subset of Sec exported proteins, contain distinct lipoprotein signal peptides with a lipobox motif, often L-A-G/A-↓C where ↓ represents the signal peptide cleavage site (Juncker *et al.*, 2003). The conserved cysteine residue immediately following the signal peptide cleavage site becomes lipid modified upon export.

Lipoproteins can be predicted using LipoP (Juncker *et al.*, 2003), and a comprehensive *in silico* analysis of lipoproteins in mycobacteria has recently been published (Sutcliffe & Harrington, 2004).

Several programs are designed to specifically predict signal peptides for the Tat pathway including Tatfind (Rose *et al.*, 2002), TatPred (Taylor *et al.*, 2006), TatP (Bendtsen *et al.*, 2005), and TigrFAM (Selengut *et al.*, 2007). These Tat prediction algorithms are built on slight variations of the twin arginine motif, commonly expressed as R-R-X-Φ-Φ (Φ = hydrophobic) (Ligon *et al.*, 2012), and some incorporate additional requirements for the signal peptide regions surrounding the Tat motif (Bendtsen *et al.*, 2005; Rose *et al.*, 2002). As previously described (McDonough *et al.*, 2008), variations between Tat prediction algorithms results in disagreement as to which *M. tuberculosis* proteins are predicted to contain a Tat signal peptide.

Programs specific to identification of membrane proteins include TMHMM (Krogh *et al.*, 2001; Sonnhammer *et al.*, 1998) TMPRED (Hofmann & Stoffel, 1993), TopPred (Claros & von Heijne, 1994), and MEMSAT2 (Jones *et al.*, 1994) which predict transmembrane domains through the identification of large regions of hydrophobic residues forming alpha-helices (Krogh *et al.*, 2001; Sonnhammer *et al.*, 1998; Hofmann & Stoffel, 1993; Jones *et al.*, 1994; Claros & von Heijne, 1994), as well as charge analysis based on positively charged residues on the cytoplasmic face of the membrane (von Heijne, 1992; Krogh *et al.*, 2001). Studies generally predict that 20-30% of proteins in most genomes contain transmembrane domains (Punta *et al.*, 2007). Transmembrane prediction algorithms which rely only on hydrophobicity analysis lead to a false identification of hydrophobic regions in globular proteins as transmembrane domains, and care should be taken when choosing a program to identify membrane proteins (Punta *et al.*, 2007). Of available programs, TMHMM was one identified as having a low incidence of false

positive prediction of globular proteins (Punta *et al.*, 2007). Consistent with previously analyzed bacterial genomes, TMHMM predicts 802 proteins (20% of all proteins) in *M. tuberculosis* to be membrane proteins (Perkowski, unpublished data). All transmembrane prediction algorithms falsely identify N-terminal signal peptides as transmembrane domains due to their inherent hydrophobicity (Punta *et al.*, 2007). Therefore, predicted transmembrane domains at the far N-terminus of proteins should be further inspected for cleavage sites to discriminate between N-terminal transmembrane domains and signal peptides.

Transmembrane prediction programs will also predict the final topology of the integral membrane proteins and the protein domains that will be localized to the cytoplasm or periplasm. Unfortunately, these predictions frequently disagree, and currently no method is accurate at predicting topology from sequence information alone (Punta *et al.*, 2007). Prediction algorithms are built based on high resolution protein structure information; however, very few membrane proteins have solved crystal structures (Punta *et al.*, 2007). Additionally, transmembrane topology may be determined not only by sequence, but also by interactions with the export machinery (Punta *et al.*, 2007). For these reasons, it is essential that *in silico* predicted topologies of transmembrane proteins be experimentally validated.

For the SecA2 and ESX protein export systems of mycobacteria there currently exist no bioinformatics prediction programs, primarily due to the fact that the precise requirements for export through these pathways are only beginning to be discovered. Of the known SecA2 exported proteins many contain signal peptides that are identified with the existing signal peptide bioinformatics prediction programs mentioned above (Gibbons *et al.*, 2007; Feltcher *et al.*, 2015). However, some proteins thought to be exported by the SecA2 pathway in *M. tuberculosis* and *M. marinum*, such as SodA and PknG, do not have canonical signal peptides (Braunstein *et*

al., 2003; van der Woude *et al.*, 2014; Feltcher *et al.*, 2015). Additionally, there are features of the mature domain of SecA2 exported proteins that are currently poorly defined and required for export, and the leading hypothesis is that the mature domain of SecA2 exported proteins tends to fold in the cytoplasm (Feltcher *et al.*, 2013). Export through ESX Type VII secretion systems requires a YxxxD/E secretion motif within the exported protein or a partner co-secreted protein (Daleke *et al.*, 2012). However, the YxxxD/E motif is not sufficient for export because undefined elements outside of this motif determine which ESX system a protein is exported through (Daleke *et al.*, 2012). Overall, while bioinformatics programs play an important role in predicting exported proteins in *M. tuberculosis*, they have limitations and should not be the only method used to determine if a given protein is exported.

Experimental Identification of Exported Proteins: Mass Spectrometry

Mass spectrometry based methods can be used to identify the proteins localized to exported fractions of *M. tuberculosis* (cytoplasmic membrane, cell wall or fully secreted culture filtrate) as recently reviewed (de Souza & Wiker, 2011). For these studies, *M. tuberculosis* is grown *in vitro* and then fractionated by various methods to isolate subcellular compartments. The power of mass spectrometry based approaches lies in the sheer quantity of proteins that are identified and the potential to acquire information about protein abundance (Solis & Cordwell, 2011; de Souza & Wiker, 2011). Quantitative mass spectrometry, in particular, is a powerful method for identifying substrates of specific export systems, by comparing the abundance of proteins in exported fractions between wild-type and export system mutants (Feltcher *et al.*, 2015; Champion *et al.*, 2014; Altindis *et al.*, 2015; Fritsch *et al.*, 2013). However, there are limitations to mass spectrometry of subcellular fractions as a means of comprehensive identification of exported proteins (Solis & Cordwell, 2011; de Souza & Wiker, 2011; Yang *et*

al., 2015; Schmidt & Volker, 2011). While enrichment of exported proteins is successfully accomplished with subcellular fractionation methods, contamination with cytoplasmic material can never be fully avoided. This often leads to the identification of highly expressed cytoplasmic protein contaminants in exported fractions (Table 1.1) (de Souza & Wiker, 2011). Furthermore, as mass spectrometry has become increasingly sensitive, the potential for false positive identification of cytoplasmic contaminants as exported proteins has grown. In recent mass spectrometry proteomics studies, only 30% of proteins identified in exported fractions have a predicted export signal, highlighting the abundance of non-exported proteins being detected (Table 1.1). Further, in one of the most comprehensive proteomics studies of *M. tuberculosis* to date, of 1,050 *M. tuberculosis* proteins identified 98% of the proteins were detected in both cytoplasmic and exported fractions (Bell *et al.*, 2012). Thus, caution should be exercised when a protein is identified as being exported solely on the basis of mass spectrometry based proteomics.

An additional challenge facing exported protein identification by mass spectrometry is the use of this technology to identify proteins exported by a pathogen during *in vivo* growth in a host (Schmidt & Volker, 2011; Yang *et al.*, 2015). Mass spectrometry based approaches are biased by abundance, and low abundance proteins in a sample are difficult to identify. This becomes particularly difficult when attempting to identify relatively rare bacterial proteins from among host cells proteins in an infection model (Schmidt & Volker, 2011; Yang *et al.*, 2015; Kruh *et al.*, 2010). Isolation and purification of bacterial cells out of the host tissue, as well as very high bacterial cell numbers is generally required (Xia *et al.*, 2007; Twine *et al.*, 2006; Becker *et al.*, 2006; Schmidt *et al.*, 2010; Liu *et al.*, 2012; Pieper *et al.*, 2009). Further, it is hard to determine if a bacterial protein identified in the context of infection was cytoplasmic or

exported. Analysis of crude extracts of infected organs does not allow for discrimination between exported and non-exported proteins (Schmidt & Volker, 2011; Yang *et al.*, 2015). Isolation of bacterial cells from infected host cells could be useful for identifying proteins exported to the bacterial cytoplasmic membrane and cell wall, but current methods will not provide sufficient material for subcellular fractionation and further analysis (Schmidt & Volker, 2011; Yang *et al.*, 2015). An alternative approach is to search for bacterial proteins secreted away from the bacteria into the host cytosol. However, due to the sensitivity of mass spectrometry this demands an efficient method to eliminate intracellular bacteria from the host cytosol which represents another technical challenge. As a result, the exported proteomes of pathogens during infection remain to be defined.

Metabolic labeling and biorthogonal chemistry promises to improve many fields, including the study of bacterial exported proteins (Ngo *et al.*, 2009; Horisawa, 2014; Mahdavi *et al.*, 2014; Siegrist *et al.*, 2015). Incorporating labels amenable to click-chemistry can allow for downstream chemical reactions which can tag labeled compounds with fluorescent markers, or allow for isolation of tagged molecules out of complex mixtures (Mahdavi *et al.*, 2014; Dieterich *et al.*, 2007; Siegrist *et al.*, 2015). Recently, bio-orthogonal noncanonical amino acid tagging, BONCAT (Dieterich *et al.*, 2007), has been used in *Escherichia coli, Salmonella enterica* serovar Typhimurium and *Yersinia pseudotuberculosis* to incorporate azidonorleucine (ANL) into proteins (Tanrikulu *et al.*, 2009; Grammel *et al.*, 2010; Mahdavi *et al.*, 2014). In one study, *S.* Typhimurium was used to infect mammalian cells *in vitro*, cells were pulsed with ANL to label all bacterial proteins, and then click chemistry was used to selectively tag bacterial ANL labeled proteins with biotin for isolation (Grammel *et al.*, 2010). A second study used the same technology to isolate proteins secreted by the type three secretion system (T3SS) of extracellular

Y. pseudotuberculosis into *in vitro* cultured cells (Mahdavi *et al.*, 2014). Exported proteins delivered into the host cell were identified through the comparison of wild type *Y. pseudotuberculosis* to a T3SS mutant unable to translocate effectors (Mahdavi *et al.*, 2014). Unfortunately, a large number of background cytoplasmic bacterial proteins were also identified in the host cell, which prevented applying this technology to the identification of new exported proteins (Mahdavi *et al.*, 2014). Overall, while this technology holds promise for facilitating studies of proteins exported in the context of infection, there remain significant technical hurdles. In addition, incorporation of unnatural amino acids by bacteria has yet to be performed during *in vivo* infection.

Experimental Identification of Exported Proteins: Genetic Reporters

An alternative strategy to biochemical identification of exported proteins is the use of genetic reporters of protein export. A protein export reporter is an enzyme whose activity or function depends on the reporter being exported/localized out of the cytoplasm. Protein export reporters are missing their native signal for export and, consequently, they require in-frame fusion with a protein capable of exporting the reporter to be active. Genetic reporters require a genetically tractable organism and they need to be compatible with the specific export system under investigation. Genetic reporter approaches for identifying exported proteins have the benefit over mass spectrometry of identifying fewer false positives and thus cytoplasmic proteins are less of a concern with export reporters. Over 90% of proteins identified with a genetic reporter contain a predicted export signal (Table 1.1, Figure 1.4), which reinforces their identification as being exported. This is compared to by mass spectrometry based methods which only identify approximately 30% of proteins with predicted export signals. Comparing studies of exported proteins in *M. tuberculosis* using mass spectrometry or genetic reporter based

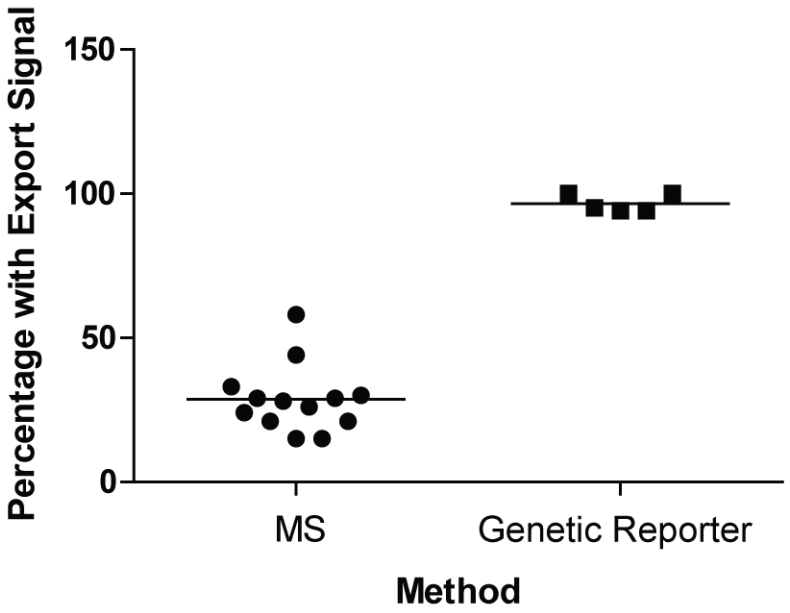

Figure 1.4. Genetic Reporters Identify a High Proportion of Proteins with Export Signals. Data calculated in Table 1.1 was graphed to demonstrate that technologies utilizing a genetic reporter result in identification of primarily exported proteins, as defined by containing an export signal. Export signals including transmembrane domains and Sec and Tat signal peptides were predicted by SignalP, TatP, TMHMM, (Sutcliffe & Harrington, 2004), and (McDonough *et al.*, 2008).

approaches reveals high abundance cytoplasmic proteins (as identified by PaxDB (Wang *et al.*, 2015; Wang *et al.*, 2012)) as frequently identified as exported by mass spectrometry but not by genetic reporter approaches (Table 1.1, Figure 1.5).

The classic genetic reporter of bacterial protein export is the enzyme alkaline phosphatase (PhoA) (Taylor *et al.*, 1987; Kaufman & Taylor, 1994; Cleavinger *et al.*, 1995; Lim *et al.*, 1995; Manoil & Beckwith, 1985; Manoil *et al.*, 1990). Bacterial colonies exporting PhoA can be detected as blue colonies when plate on agar containing a colorimetric PhoA substrate. Consequently, a truncated 'PhoA reporter lacking its native signal for export can "report" on the presence of an export signal in a protein to which it is fused, by producing blue colored colonies. In the first study to identify *M. tuberculosis* exported proteins using a 'PhoA reporter, 'PhoA was encoded on a plasmid, and random fragments of *M. tuberculosis* genomic DNA were fused upstream of the reporter (Lim *et al.*, 1995). The resulting plasmids were screened in *Mycobacterium smegmatis*, a non-pathogenic model species of *M. tuberculosis*, and blue colonies were identified when the transformants were plated on the PhoA colorimetric substrate (Lim *et al.*, 1995). This first study identified three *M. tuberculosis* proteins as exported (Lim *et al.*, 1995). Further studies utilized the 'PhoA reporter combined with *in vitro* transposition into *M. tuberculosis* cosmids. These transposon mutagenized cosmids were transformed into *M. smegmatis* to screen for active PhoA fusions, resulting in the identification of 31 *M. tuberculosis* exported proteins (Braunstein *et al.*, 2000). Unfortunately, due to endogenous phosphatase activity, the 'PhoA reporter is not compatible with use in *M. tuberculosis*, and thus all 'PhoA screens have to be carried out in the non-pathogenic *M. smegmatis* (Lim *et al.*, 1995).

Generation of a β-lactam sensitive *ΔblaC* mutant of *M. tuberculosis* (Flores *et al.*, 2005; McDonough *et al.*, 2005) opened up the possibility of using β-lactamase reporters to identify

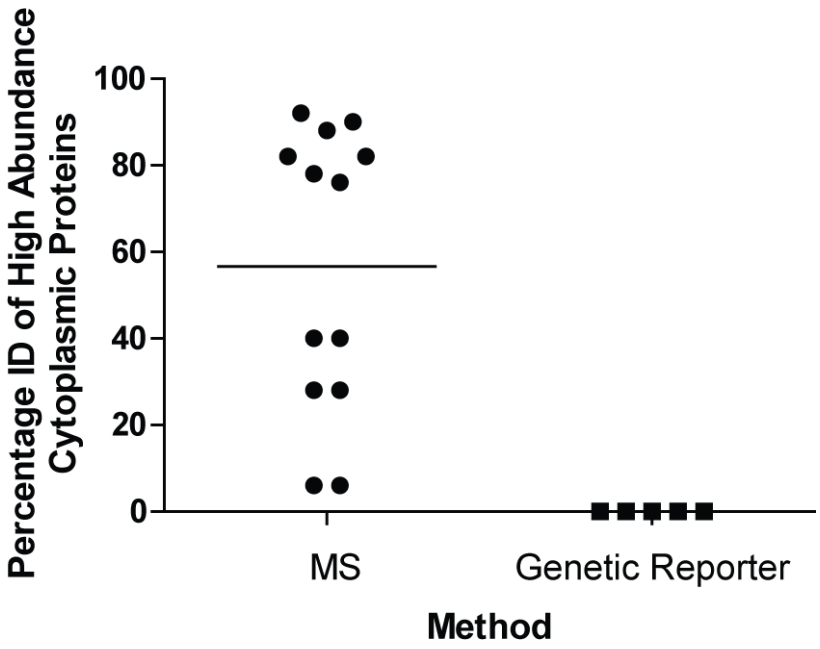

Figure 1.5. MS based methods Identify (ID) High Abundance Cytoplasmic Proteins as Exported. Percentage of high abundance cytoplasmic proteins identified in exported fractions as described in Table 1.1. The Top 50 most abundant cytoplasmic proteins were identified by PaxDB (Wang *et al.*, 2015; Wang *et al.*, 2012). Proteins identified as abundant in PaxDB but containing predicted export signals, or known to be exported (i.e. EsxA and EsxB) were excluded, to generate a list of the top 50 most abundant cytoplasmic proteins in *M. tuberculosis*.

exported proteins directly in pathogenic *M. tuberculosis*. The *E. coli* β-lactamase, BlaTEM, cleaves β-lactam antibiotics and is able to produce β-lactam resistance. Like ‘PhoA, removal of the native signal peptide of ‘BlaTEM prevents export but when fused to a signal peptide, or the extracytoplasmic portion of an exported protein, the exported ‘Bla reporter confers β-lactam resistance (Figure 1.6). This was first worked out in *Escherichia coli* (Broome-Smith & Spratt, 1986), and holds true when the ‘BlaTEM reporter is used in the *ΔblaC* mutant of *M. tuberculosis* (McCann *et al.*, 2007). The ‘BlaTEM reporter is compatible with both Sec and Tat export systems of *M. tuberculosis* (McCann *et al.*, 2007). An additional β-lactamase reporter (‘BlaC) has been developed for *M. tuberculosis* that is specific for export by the Tat pathway (McDonough *et al.*, 2005; McDonough *et al.*, 2008). For the ‘BlaC reporter, only fusion to Tat exported proteins (and not signals for other export systems) works to confer resistance to β-lactams (McDonough *et al.*, 2008). Generation of a library of plasmids containing the *‘blaC* reporter fused downstream of random fragments of *M. tuberculosis* genomic DNA led to the identification of 13 proteins exported by the Tat pathway in *M. tuberculosis* (McDonough *et al.*, 2008). The ‘*blaC* reporter has additionally been used to directly test individual proteins for Tat dependent export (Ligon *et al.*, 2012).

In other bacterial pathogens, transposons carrying the ‘PhoA reporter proved a powerful means to identify exported proteins with roles in virulence (Taylor *et al.*, 1987; Kaufman & Taylor, 1994; Manoil & Beckwith, 1985; Manoil *et al.*, 1990). With the development of β-lactamase reporters that work in mycobacteria, this approach could be directly applied to the identification of exported virulence factors of *M. tuberculosis* Using a *Himar1*-based mariner transposon carrying the ‘BlaTEM reporter in *M. tuberculosis* and plating on β-lactam containing agar, transposon insertions in genes encoding 111 unique *M. tuberculosis* exported proteins were

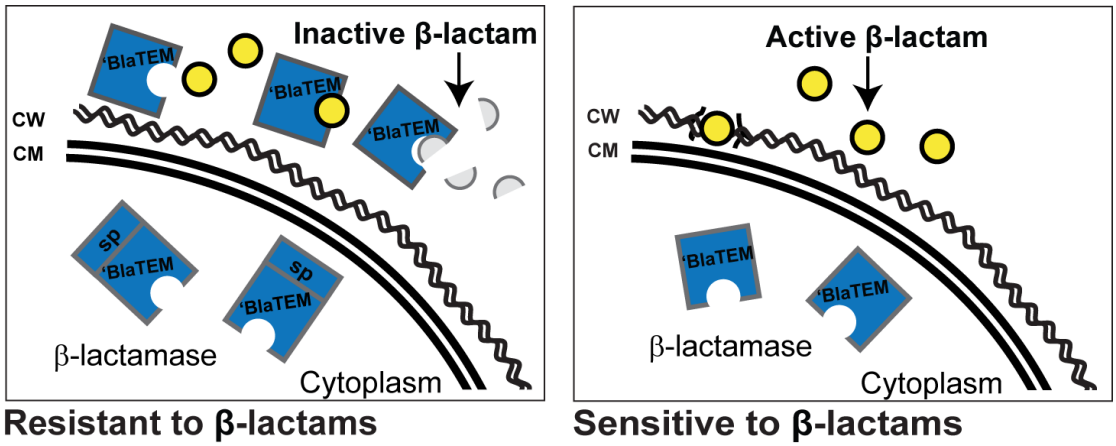

Figure 1.6. 'BlaTEM Reporter Activity is Dependent on Export. The β-lactamase BlaTEM from *E. coli* (blue) contains a native signal peptide that directs it to be exported out of the cytoplasm, where it breaks down β-lactam antibiotics (yellow) and confers β-lactam resistance (left). The 'BlaTEM reporter is composed of a truncated protein where the native signal peptide has been removed, and thus the reporter is not exported. This results in β-lactam sensitivity, when β-lactam antibiotics break down the cell wall (right). Fusion of the reporter to a signal peptide or the extra-cytoplasmic portion of an exported protein will rescue export and confer resistance to β-lactam antibiotics (left).

identified (McCann *et al.*, 2011). The same in-frame transposon insertion that identifies an exported protein will, in most cases, also disrupt the function of that exported protein. By subsequent screening of 111 transposon insertion mutants in genes encoding exported proteins for virulence defects in a macrophage intracellular growth assay a total of six exported proteins with roles in virulence were identified (McCann *et al.*, 2011). One of these proteins, Rv0199, had no previously predicted function. Chapter 3 will focus on further characterization of the *rv0199* transposon mutant culminating in assigning function to Rv0199, and shedding light on the function of a previously uncharacterized exported protein family.

There are advantages to using a β-lactamase based genetic reporter over other genetic reporters. First, β-lactamases are selectable markers, dramatically reducing the otherwise laborious nature of screening for reporter positive colonies (McCann *et al.*, 2011). Second, β-lactamases have the potential of overcoming the requirement for using *in vitro* grown bacteria when identifying exported proteins. Many β-lactam antibiotics are available and safe for use in cell culture and animal models. Proof of principle studies demonstrated, that the *ΔblaC M. tuberculosis* mutant is susceptible to β-lactams during intracellular growth in a cultured macrophage cell line (McCann *et al.*, 2007). Furthermore, McCann *et al.* demonstrated that the ‘BlaTEM reporter can report on the export of a protein to which it is fused during intracellular growth. Model strains were constructed that produce the ‘BlaTEM reporter with no export signal (‘BlaTEM, β-lactam sensitive) or the ‘BlaTEM reporter fused to a signal peptide (sp-‘BlaTEM, β-lactam resistant) in *M. tuberculosis* (McCann *et al.*, 2007). Macrophages were infected with each strain, and treated by addition of the β-lactam carbenicillin to the media. In the absence of treatment both strains could replicate in macrophages; however, carbenicillin treatment restricted the intracellular growth of the ‘BlaTEM β-lactam sensitive strain. The strain with an exported sp-

‘BlaTEM reporter was protected from carbenicillin treatment, and grew normally in the macrophage. This demonstrated that the ‘BlaTEM reporter can report on export during growth in macrophages (McCann *et al.*, 2007). These studies with macrophages, suggested the possibility of using the ‘BlaTEM reporter in screening for *M. tuberculosis* proteins exported during infection in β-lactam treated mice. Chapter 2 will focus on the development of a method called EXIT (EXported *In vivo* Technology) which we developed to specifically identify proteins exported by *M. tuberculosis* during murine infection.

Exported proteins play important roles in *M. tuberculosis* virulence

Exported proteins are vital to the survival of intracellular pathogens, such as *M. tuberculosis*, where they not only maintain cellular physiology, but also provide protection from intracellular stresses, control the host cell response, and allow for nutrient acquisition in a purposefully nutrient poor environment (McCann, 2009). Several examples of exported virulence factors are highlighted below.

Exported proteins important to M. tuberculosis virulence

M. tuberculosis encodes several classical exported virulence factors, for example, proteins involved in detoxification of radical oxygen and radical nitrogen species (McCann, 2009). Additionally, some of the best studied virulence factors in *M. tuberculosis* are exported proteins required for the biosynthesis specialized lipids within the highly complex and impermeable mycobacterial cell wall and outer membrane (myco-membrane). The myco-membrane consists of several specialized lipid structures, many including long chain fatty acids called mycolic acids. Synthesis and transport of many lipids within the myco-membrane of *M. tuberculosis* is required for bacterial cell viability; however, some lipids are not required for *in vitro* growth, including phthiocerol dimycocerosate (PDIM). Mutations in genes required for the

synthesis and transport of PDIM to the myco-membrane results in attenuation of *M. tuberculosis* during infection of macrophages and mice (Forrellad *et al.*, 2013). Interestingly, PDIM appears to be important for immune modulation, and the attenuation of PDIM mutants appears to result from reduced control of the host (Forrellad *et al.*, 2013).

M. tuberculosis has not only had to develop strategies to evade destruction by the host immune system, but also to acquire nutrients for survival. The phagosome is nutritionally restricted as a way to prevent growth of intracellular bacteria (Appelberg, 2006). Perhaps the most well studied nutritional restriction, the host's ability to restrict iron is overcome by *M. tuberculosis* through the production of siderophores, molecules with very high affinity for iron which are able to chelate it away from host iron stores (Fang *et al.*, 2015). Disruption of siderophore biosynthesis or function attenuates the virulence of *M. tuberculosis* in cell culture and animal models of infection (Fang *et al.*, 2015).

Role of Lipid Import and Catabolism in M. tuberculosis virulence

During infection, *M. tuberculosis* is thought to primarily consume lipids as a carbon source (Munoz-Elias & McKinney, 2006; Eisenreich *et al.*, 2010). This lifestyle is reflected in the numerous enzymes required for catabolism of lipids in *M. tuberculosis* (Cole *et al.*, 1998). Additionally, the *M. tuberculosis* genome encodes several specialized multi-protein transporters predicted to import lipids, four Mce transporters (Mce1, 2, 3, and 4). All of the *mce* operons are important to virulence in mice (Sassetti & Rubin, 2003; Shimono *et al.*, 2003; Lima *et al.*, 2007; Uchida *et al.*, 2007; Senaratne *et al.*, 2008; Marjanovic *et al.*, 2010). Study of *mce* operons in *M. tuberculosis* has shown that one of them, *mce4*, imports cholesterol (Pandey & Sassetti, 2008; Klepp *et al.*, 2012; Mohn *et al.*, 2008). *M. tuberculosis* is fully able to catabolize cholesterol (Miner *et al.*, 2009), and mutants have been generated in several steps of the cholesterol

metabolic pathway, and they are severely attenuated (Miner *et al.*, 2009). The *mce4* mutant is attenuated in a mouse model of infection, as demonstrated by reduced bacterial burden, increased survival time in mice, and decreased histopathology, demonstrating an importance for cholesterol import during infection (Senaratne *et al.*, 2008; Pandey & Sassetti, 2008). The other *mce* operons are not as well studied as *mce4*, but they are also thought to be lipid transporters. Mce1 is required during infection of macrophages as well as during murine infection (McCann *et al.*, 2011; Rengarajan *et al.*, 2005). Recent evidence suggests that Mce1 imports mycolic acids, specialized lipids which are incorporated into many complex molecules in the myco-membrane (Forrellad *et al.*, 2014; Cantrell *et al.*, 2013). Mce1 functions potentially as a mechanism of recycling for control of the mycolic acid concentration within the mycobacterial outer membrane. The lipids transported by Mce2 or Mce3 have yet to be identified.

Although the contribution of Mce systems to *M. tuberculosis* virulence has been characterized, the role of individual proteins within the Mce transport complex has yet to be characterized. *Mce* operons are thought to be analogous to ABC transporters, and the genes encoded in the operons have been assigned function based on our knowledge of ABC transporters (Casali & Riley, 2007). However, the sequence similarity to ABC transporters is limited, and Mce transporters are distinct from ABC transporters in that *mce* operons contain genes encoding multiple predicted permease and solute binding protein components. *mce* operons additionally contain genes encoding proteins with no homology to ABC transporter components that have been named Mce-associated (Mas) proteins. Therefore, the mechanism of Mce transport and the function of the individual Mce transporter components in lipid transport and virulence await direct study. In Chapter 3, we characterize Rv0199 (renamed OmasA) as an

in vivo exported protein of unknown function and assign it a role in the stabilization and function of Mce transporter systems.

Summary

While the technology for identifying exported proteins has improved dramatically, there are still technical hurdles to accurate genome-wide identification of exported proteins. Furthermore, current technologies remain limited to identifying exported proteins using *in vitro* grown bacteria. Transcriptional analysis has identified large shifts in the *M. tuberculosis* lifestyle during infection (Talaat *et al.*, 2004; Schnappinger *et al.*, 2006). While transcriptional analysis has informed our perspectives on *M. tuberculosis* metabolism, and the intracellular lifestyle, there is mounting evidence that transcriptional analysis alone is not representative of cellular protein changes (de Souza & Wiker, 2011). Thus, it is important to develop methods to specifically study protein behavior during infection.

We propose that a subset of exported proteins may be selectively expressed or exported during infection, that are missed by current analyses focusing on *in vitro* grown *M. tuberculosis*. Such proteins could be incredibly important to pathogenesis, and define interactions between the host and pathogen. In this dissertation, we sought to comprehensively identify the *in vivo* exported proteome of *M. tuberculosis*. Accomplishment of this goal, which has not been attempted previously in any bacterial pathogen, required the development of a novel method, which we termed EXIT for EXported *In vivo* Technology. Chapter 2 will describe the process of developing and optimizing EXIT, as well as describing key results from analyzing the *in vivo* exported proteome of *M. tuberculosis*. We identified 593 proteins as exported during murine infection, including 32 proteins with no *in silico* predicted export signal which represent candidates for new exported proteins. We further identified 38 proteins whose export was

significantly upregulated during infection as compared to *in vitro* growth. Because these proteins are subject to regulation, they may have important functions during infection and could understand the host-pathogen interface. Finally, we applied our reporter fusion data to determine the topology of important *M. tuberculosis* virulence factors. The proteins identified in this study, and the vast resource of topology data, represent a valuable resource for the *M. tuberculosis* research community.

Identification of exported proteins is a proximal goal. With the knowledge that exported proteins are critical to *M. tuberculosis* virulence we also seek to better understand the roles of individual exported proteins during infection. Previous studies analyzing transposon mutants lacking individual exported proteins during intracellular growth identified six attenuated mutants (McCann *et al.*, 2011). Of these, one mutation was in a membrane protein of unknown function, Rv0199. Rv0199 was also identified by EXIT as exported *in vivo*. Chapter 3 describes further characterization of the *rv0199* mutant, resulting in a functional assignment of *rv0199* as an *orphaned mce-associated* gene (*mas*). Further study into the role of Rv0199 in Mce transport determined that Rv0199, and potentially all Mas family proteins, provides stability to the large multi-protein Mce transport complex.

Table 1.1Exported Protein Identification in *M. tuberculosis*						
First Author	**Year**	**Method**	**Fraction**	**# ID Proteins**	**% Export Signal**	**% ID of Top 50 Highly Expressed Cytoplasmic Proteins**
Gu	2003	MS	MEM	737	15	82
Rosenkrands	2000	MS	CF	49	29	28
Xiong	2005	MS	MEM	342	44	40
Mawuenyega	2005	MS	MEM	105	21	6
Mawuenyega	2005	MS	CW	65	15	6
Wolfe	2010	MS	CW	540	26	76
Malen	2007	MS	MEM	255	58	28
de Souza	2011	MS	CF	458	33	40
de Souza	2011	MS	MEM	1439	30	90
Bell	2012	MS	CW	791	21	92
Bell	2012	MS	MEM	666	24	78
Bell	2012	MS	CF	508	29	82
Gunawardena	2013	MS	MEM	2203	28	88
Feltcher	2015	MS	CW	1729	29	82
Braunstein	2000	GR	Exported	31	94	0
Gomez	2000	GR	Exported	9	100	0
McDonough	2008	GR	Exported	13	100	0
McCann	2011	GR	Exported	111	95	0

Table 1.1. Exported Protein Identification in *M. tuberculosis*. Exported proteins of *M. tuberculosis* have been identified by mass-spectrometry (MS) based methods as well as use of genetic reporters (GR) (Gu *et al.*, 2003; Rosenkrands *et al.*, 2000; Xiong *et al.*, 2005; Mawuenyega *et al.*, 2005; Wolfe *et al.*, 2010; Malen *et al.*, 2007; de Souza *et al.*, 2011; Bell *et al.*, 2012; Gunawardena *et al.*, 2013; Braunstein *et al.*, 2000; Gomez *et al.*, 2000; McDonough *et al.*, 2008; McCann *et al.*, 2011). MS based methods have analyzed proteins in exported fractions including the membrane (MEM), cell wall (CW), and extracellular or culture filtrate (CF) fractions. For comparison, all lists were analyzed using the H37Rv RefSeq genome annotation released January 9 2012. The total number of proteins identified (# ID Proteins) in the given fraction is reported and analyzed for percentage containing an export signal (% Export Signal) and the percentage of top 50 most abundant cytoplasmic proteins identified (% ID Top Cytoplasmic), as described below. Export signals including transmembrane domains and Sec and Tat signal peptides were predicted by SignalP, TatP, TMHMM, (Sutcliffe & Harrington, 2004), and (McDonough *et al.*, 2008). The Top 50 most abundant cytoplasmic proteins were identified by PaxDB, a database calculating the relative abundance of proteins from published proteomics datasets (Wang *et al.*, 2015; Wang *et al.*, 2012). Proteins identified as abundant in PaxDB but containing predicted export signals, or known to be exported (i.e. EsxA and EsxB) were excluded, to generate a list of the top 50 most abundant cytoplasmic proteins in *M. tuberculosis*.

REFERENCES

Genome analysis: comparison of the transport capabilities of several bacteria. *http://www.biology.ucsd.edu/~ipaulsen/transport/*.

Abdallah, A.M., T. Verboom, E.M. Weerdenburg, N.C. Gey van Pittius, P.W. Mahasha, C. Jimenez, M. Parra, N. Cadieux, M.J. Brennan, B.J. Appelmelk & W. Bitter, (2009) PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. *Mol Microbiol* **73**: 329-340.

Agbor, T.A. & B.A. McCormick, (2011) Salmonella effectors: important players modulating host cell function during infection. *Cell Microbiol* **13**: 1858-1869.

Altindis, E., T. Dong, C. Catalano & J. Mekalanos, (2015) Secretome analysis of *Vibrio cholerae* type VI secretion system reveals a new effector-immunity pair. *mBio* **6**: e00075.

Appelberg, R., (2006) Macrophage nutriprive antimicrobial mechanisms. *J Leukoc Biol* **79**: 1117-1128.

Armstrong, J.A. & P.D. Hart, (1971) Response of cultured macrophages to *Mycobacterium tuberculosis*, with observations on fusion of lysosomes with phagosomes. *J Exp Med* **134**: 713-740.

Banu, S., N. Honore, B. Saint-Joanis, D. Philpott, M.C. Prevost & S.T. Cole, (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol. Micro* **44**: 9-19.

Barry, C.E., 3rd, H.I. Boshoff, V. Dartois, T. Dick, S. Ehrt, J. Flynn, D. Schnappinger, R.J. Wilkinson & D. Young, (2009) The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol* **7**: 845-855.

Becker, D., M. Selbach, C. Rollenhagen, M. Ballmaier, T.F. Meyer, M. Mann & D. Bumann, (2006) Robust Salmonella metabolism limits possibilities for new antimicrobials. *Nature* **440**: 303-307.

Bell, C., G.T. Smith, M.J. Sweredoski & S. Hess, (2012) Characterization of the *Mycobacterium tuberculosis* proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. *J Proteome Res* **11**: 119-130.

Bendtsen, J.D., H. Nielsen, D. Widdick, T. Palmer & S. Brunak, (2005) Prediction of twin-arginine signal peptides. *BMC Bioinformatics* **6**: 167.

Bottai, D., M. Di Luca, L. Majlessi, W. Frigui, R. Simeone, F. Sayes, W. Bitter, M.J. Brennan, C. Leclerc, G. Batoni, M. Campa, R. Brosch & S. Esin, (2012) Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. *Mol Microbiol* **83**: 1195-1209.

Braunstein, M., A.M. Brown, S. Kurtz & W.R. Jacobs, Jr., (2001) Two nonredundant SecA homologues function in mycobacteria. *J Bacteriol* **183**: 6979-6990.

Braunstein, M., B. Espinosa, J. Chan, J.T. Belisle & W.R.J. Jacobs, (2003) SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. *Molecular Microbiology* **48**: 453-464.

Braunstein, M., T.I. Griffin, J.I. Kriakov, S.T. Friedman, N.D. Grindley & W.R. Jacobs, Jr., (2000) Identification of genes encoding exported *Mycobacterium tuberculosis* proteins using a Tn552'phoA in vitro transposition system. *J Bacteriol* **182**: 2732-2740.

Brennan, M.J., G. Delogu, Y. Chen, S. Bardarov, J. Kriakov, M. Alavi & W.R. Jacobs, Jr., (2001) Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun* **69**: 7326-7333.

Broome-Smith, J.K. & B.G. Spratt, (1986) A vector for the construction of translational fusions to TEM beta-lactamase and the analysis of protein export signals and membrane protein topology. *Gene* **49**: 341-349.

Cantrell, S.A., M.D. Leavell, O. Marjanovic, A.T. Iavarone, J.A. Leary & L.W. Riley, (2013) Free mycolic acid accumulation in the cell wall of the mce1 operon mutant strain of *Mycobacterium tuberculosis*. *Journal of microbiology* **51**: 619-626.

Casali, N. & L.W. Riley, (2007) A phylogenomic analysis of the Actinomycetales mce operons. *BMC Genomics* **8**: 60.

Champion, M.M., E.A. Williams, R.S. Pinapati & P.A. Champion, (2014) Correlation of phenotypic profiles using targeted proteomics identifies mycobacterial esx-1 substrates. *J Proteome Res* **13**: 5151-5164.

Claros, M.G. & G. von Heijne, (1994) TopPred II: an improved software for membrane protein structure predictions. *Computer applications in the biosciences : CABIOS* **10**: 685-686.

Cleavinger, C.M., M.F. Kim, J.H. Im & K.S. Wise, (1995) Identification of mycoplasma membrane proteins by systematic Tn *phoA* mutagenesis of a recombinant library. *Mol Microbiol* **18**: 283-293.

Cole, S.T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels & B.G. Barrell, (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

Daleke, M.H., R. Ummels, P. Bawono, J. Heringa, C.M. Vandenbroucke-Grauls, J. Luirink & W. Bitter, (2012) General secretion signal for the mycobacterial type VII secretion pathway. *Proc Natl Acad Sci U S A* **109**: 11342-11347.

Daniel, T.M., (2006) The history of tuberculosis. *Respiratory medicine* **100**: 1862-1870.

de Souza, G.A., N.A. Leversen, H. Malen & H.G. Wiker, (2011) Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. *Journal of proteomics* **75**: 502-510.

de Souza, G.A. & H.G. Wiker, (2011) A proteomic view of mycobacteria. *Proteomics* **11**: 3118-3127.

Dieterich, D.C., J.J. Lee, A.J. Link, J. Graumann, D.A. Tirrell & E.M. Schuman, (2007) Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat Protoc* **2**: 532-540.

Eisenreich, W., T. Dandekar, J. Heesemann & W. Goebel, (2010) Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nat Rev Microbiol* **8**: 401-412.

Facey, S.J. & A. Kuhn, (2010) Biogenesis of bacterial inner-membrane proteins. *Cell Mol Life Sci* **67**: 2343-2362.

Fang, Z., S.L. Sampson, R.M. Warren, N.C. Gey van Pittius & M. Newton-Foot, (2015) Iron acquisition strategies in mycobacteria. *Tuberculosis (Edinb)*.

Feltcher, M.E. & M. Braunstein, (2012) Emerging themes in SecA2-mediated protein export. *Nat Rev Microbiol* **10**: 779-789.

Feltcher, M.E., H.S. Gibbons, L.S. Ligon & M. Braunstein, (2013) Protein export by the mycobacterial SecA2 system is determined by the preprotein mature domain. *J Bacteriol* **195**: 672-681.

Feltcher, M.E., H.P. Gunawardena, K.E. Zulauf, S. Malik, J.E. Griffin, C.M. Sassetti, X. Chen & M. Braunstein, (2015) Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall. *Molecular & cellular proteomics : MCP*.

Flores, A.R., L.M. Parsons & M.S. Pavelka, Jr., (2005) Genetic analysis of the beta-lactamases of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* and susceptibility to beta-lactam antibiotics. *Microbiology* **151**: 521-532.

Flynn, J.L. & J. Chan, (2001) Tuberculosis: latency and reactivation. *Infect Immun* **69**: 4195-4201.

Forrellad, M.A., L.I. Klepp, A. Gioffre, J. Sabio y Garcia, H.R. Morbidoni, M. de la Paz Santangelo, A.A. Cataldi & F. Bigi, (2013) Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* **4**: 3-66.

Forrellad, M.A., M. McNeil, L. Santangelo Mde, F.C. Blanco, E. Garcia, L.I. Klepp, J. Huff, M. Niederweis, M. Jackson & F. Bigi, (2014) Role of the Mce1 transporter in the lipid homeostasis of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **94**: 170-177.

Fritsch, M.J., K. Trunk, J.A. Diniz, M. Guo, M. Trost & S.J. Coulthurst, (2013) Proteomic identification of novel secreted antibacterial toxins of the *Serratia marcescens* type VI secretion system. *Molecular & cellular proteomics : MCP* **12**: 2735-2749.

Gibbons, H.S., F. Wolschendorf, M. Abshire, M. Niederweis & M. Braunstein, (2007) Identification of two *Mycobacterium smegmatis* lipoproteins exported by a SecA2-dependent pathway. *J Bacteriol* **189**: 5090-5100.

Goldberg, M., N.K. Saini & S.A. Porcelli, (2014) *Evasion of Innate and Adaptive Immunity by Mycobacterium tuberculosis,* p. 824. ASM Press.

Gomez, M., S. Johnson & M.L. Gennaro, (2000) Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. *Infect Immun* **68**: 2323-2327.

Grammel, M., M.M. Zhang & H.C. Hang, (2010) Orthogonal alkynyl amino acid reporter for selective labeling of bacterial proteomes during infection. *Angew Chem Int Ed Engl* **49**: 5970-5974.

Gu, S., J. Chen, K.M. Dobos, E.M. Bradbury, J.T. Belisle & X. Chen, (2003) Comprehensive Proteomic Profiling of the Membrane Constituents of a *Mycobacterium tuberculosis* Strain. *Molecular & cellular proteomics : MCP* **2**: 1284-1296.

Gunawardena, H.P., M.E. Feltcher, J.A. Wrobel, S. Gu, M. Braunstein & X. Chen, (2013) Comparison of the membrane proteome of virulent *Mycobacterium tuberculosis* and the attenuated *Mycobacterium bovis* BCG vaccine strain by label-free quantitative proteomics. *J Proteome Res* **12**: 5463-5474.

Hiller, K., A. Grote, M. Scheer, R. Munch & D. Jahn, (2004) PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res* **32**: W375-379.

Hofmann, K. & W. Stoffel, (1993) TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374,**: 166-170.

Horisawa, K., (2014) Specific and quantitative labeling of biomolecules using click chemistry. *Frontiers in physiology* **5**: 457.

Houben, E.N., K.V. Korotkov & W. Bitter, (2014) Take five - Type VII secretion systems of Mycobacteria. *Biochim Biophys Acta* **1843**: 1707-1716.

Isaac, D.T. & R. Isberg, (2014) Master manipulators: an update on *Legionella pneumophila* Icm/Dot translocated substrates and their host targets. *Future Microbiol* **9**: 343-359.

Jones, D.T., W.R. Taylor & J.M. Thornton, (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **33**: 3038-3049.

Juncker, A.S., H. Willenbrock, G. Von Heijne, S. Brunak, H. Nielsen & A. Krogh, (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein science : a publication of the Protein Society* **12**: 1652-1662.

Kaufman, M.R. & R.K. Taylor, (1994) Identification of bacterial cell-surface virulence determinants with TnphoA. *Methods Enzymol* **235**: 426-448.

Klepp, L.I., M.A. Forrellad, A.V. Osella, F.C. Blanco, E.J. Stella, M.V. Bianco, L. Santangelo Mde, C. Sassetti, M. Jackson, A.A. Cataldi, F. Bigi & H.R. Morbidoni, (2012) Impact of the deletion of the six mce operons in *Mycobacterium smegmatis*. *Microbes Infect* **14**: 590-599.

Koch, R., (1932) Die aetiologie der tuberculose, a translation by Berna Pinner and Max Pinner with an introduction by Allen K. Krause. *Am Rev Tuberc* **25**: pp. 285-323.

Krogh, A., B. Larsson, G. von Heijne & E.L. Sonnhammer, (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.

Kruh, N.A., J. Troudt, A. Izzo, J. Prenni & K.M. Dobos, (2010) Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome in vivo. *PLoS One* **5**: e13938.

Kurtz, S., K.P. McKinnon, M.S. Runge, J.P. Ting & M. Braunstein, (2006) The SecA2 secretion factor of *Mycobacterium tuberculosis* promotes growth in macrophages and inhibits the host immune response. *Infect Immun*.

Leversen, N.A., G.A. de Souza, H. Malen, S. Prasad, I. Jonassen & H.G. Wiker, (2009) Evaluation of signal peptide prediction algorithms for identification of mycobacterial signal peptides using sequence data from proteomic methods. *Microbiology* **155**: 2375-2383.

Lewis, K.N., R. Liao, K.M. Guinn, M.J. Hickey, S. Smith, M.A. Behr & D.R. Sherman, (2003) Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. *J Infect Dis* **187**: 117-123.

Ligon, L.S., J.D. Hayden & M. Braunstein, (2012) The ins and outs of *Mycobacterium tuberculosis* protein export. *Tuberculosis (Edinb)* **92**: 121-132.

Ligon, L.S., N.W. Rigel, A. Romanchuk, C.D. Jones & M. Braunstein, (2013) Suppressor analysis reveals a role for SecY in the SecA2-dependent protein export pathway of Mycobacteria. *J Bacteriol* **195**: 4456-4465.

Lim, E.M., J. Rauzier, J. Timm, G. Torrea, A. Murray, B. Gicquel & D. Portnoi, (1995) Identification of *Mycobacterium tuberculosis* DNA sequences encoding exported proteins by using *phoA* gene fusions. *J Bacteriol* **177**: 59-65.

Lima, P., B. Sidders, L. Morici, R. Reader, R. Senaratne, N. Casali & L.W. Riley, (2007) Enhanced mortality despite control of lung infection in mice aerogenically infected with a *Mycobacterium tuberculosis mce1* operon mutant. *Microbes Infect* **9**: 1285-1290.

Liu, X., B. Gao, V. Novik & J.E. Galan, (2012) Quantitative Proteomics of Intracellular Campylobacter jejuni Reveals Metabolic Reprogramming. *PLoS Pathog* **8**: e1002562.

Mahdavi, A., J. Szychowski, J.T. Ngo, M.J. Sweredoski, R.L. Graham, S. Hess, O. Schneewind, S.K. Mazmanian & D.A. Tirrell, (2014) Identification of secreted bacterial proteins by noncanonical amino acid tagging. *Proc Natl Acad Sci U S A* **111**: 433-438.

Malen, H., F.S. Berven, K.E. Fladmark & H.G. Wiker, (2007) Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics* **7**: 1702-1718.

Manoil, C. & J. Beckwith, (1985) TnphoA: a transposon probe for protein export signals. *Proc Natl Acad Sci U S A* **82**: 8129-8133.

Manoil, C., J.J. Mekalanos & J. Beckwith, (1990) Alkaline phosphatase fusions: sensors of subcellular location. *J Bacteriol* **172**: 515-518.

Marchand, C.H., C. Salmeron, R. Bou Raad, X. Meniche, M. Chami, M. Masi, D. Blanot, M. Daffe, M. Tropis, E. Huc, P. Le Marechal, P. Decottignies & N. Bayan, (2012) Biochemical disclosure of the mycolate outer membrane of *Corynebacterium glutamicum*. *J Bacteriol* **194**: 587-597.

Marjanovic, O., T. Miyata, A. Goodridge, L.V. Kendall & L.W. Riley, (2010) Mce2 operon mutant strain of *Mycobacterium tuberculosis is* attenuated in C57BL/6 mice. *Tuberculosis (Edinb)* **90**: 50-56.

Mawuenyega, K.G., C.V. Forst, K.M. Dobos, J.T. Belisle, J. Chen, E.M. Bradbury, A.R. Bradbury & X. Chen, (2005) *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. *Mol Biol Cell* **16**: 396-404.

McCann, J.R., J.A. McDonough, M.S. Pavelka & M. Braunstein, (2007) Beta-lactamase can function as a reporter of bacterial protein export during *Mycobacterium tuberculosis* infection of host cells. *Microbiology* **153**: 3350-3359.

McCann, J.R., J.A. McDonough, J.T. Sullivan, M.E. Feltcher & M. Braunstein, (2011) Genome-wide identification of *Mycobacterium tuberculosis* exported proteins with roles in intracellular growth. *J Bacteriol* **193**: 854-861.

McCann, J.R., S. Kurtz & M. Braunstein, (2009) Secreted and exported proteins important to *Mycobacterium tuberculosis* pathogenesis. In: Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis. K. Wooldridge (ed). Norfolk, UK: Caister Academic Press, pp. 265-298.

McDonough, J.A., K.E. Hacker, A.R. Flores, M.S. Pavelka, Jr. & M. Braunstein, (2005) The twin-arginine translocation pathway of *Mycobacterium smegmatis* is functional and required for the export of mycobacterial beta-lactamases. *J Bacteriol* **187**: 7667-7679.

McDonough, J.A., J.R. McCann, E.M. Tekippe, J.S. Silverman, N.W. Rigel & M. Braunstein, (2008) Identification of functional Tat signal sequences in *Mycobacterium tuberculosis* proteins. *J Bacteriol* **190**: 6428-6438.

Miner, M.D., J.C. Chang, A.K. Pandey, C.M. Sassetti & D.R. Sherman, (2009) Role of cholesterol in *Mycobacterium tuberculosis* infection. *Indian J Exp Biol* **47**: 407-411.

Mohn, W.W., R. van der Geize, G.R. Stewart, S. Okamoto, J. Liu, L. Dijkhuizen & L.D. Eltis, (2008) The actinobacterial mce4 locus encodes a steroid transporter. *J Biol Chem* **283**: 35368-35374.

Munoz-Elias, E.J. & J.D. McKinney, (2006) Carbon metabolism of intracellular bacteria. *Cell Microbiol* **8**: 10-22.

Nakai, K. & P. Horton, (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**: 34-36.

Ngo, J.T., J.A. Champion, A. Mahdavi, I.C. Tanrikulu, K.E. Beatty, R.E. Connor, T.H. Yoo, D.C. Dieterich, E.M. Schuman & D.A. Tirrell, (2009) Cell-selective metabolic labeling of proteins. *Nature chemical biology* **5**: 715-717.

Orme, I.M. & R.J. Basaraba, (2014) The formation of the granuloma in tuberculosis infection. *Seminars in immunology* **26**: 601-609.

Pandey, A.K. & C.M. Sassetti, (2008) Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A* **105**: 4376-4380.

Petersen, T.N., S. Brunak, G. von Heijne & H. Nielsen, (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**: 785-786.

Pieper, R., Q. Zhang, P.P. Parmar, S.T. Huang, D.J. Clark, H. Alami, A. Donohue-Rolfe, R.D. Fleischmann, S.N. Peterson & S. Tzipori, (2009) The *Shigella dysenteriae* serotype 1 proteome, profiled in the host intestinal environment, reveals major metabolic modifications and increased expression of invasive proteins. *Proteomics* **9**: 5029-5045.

Punta, M., L.R. Forrest, H. Bigelow, A. Kernytsky, J. Liu & B. Rost, (2007) Membrane protein prediction methods. *Methods* **41**: 460-474.

Pym, A.S., P. Brodin, R. Brosch, M. Huerre & S.T. Cole, (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol Microbiol* **46**: 709-717.

Rengarajan, J., B.R. Bloom & E.J. Rubin, (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* **102**: 8327-8332.

Rohde, K., R.M. Yates, G.E. Purdy & D.G. Russell, (2007) *Mycobacterium tuberculosis* and the environment within the phagosome. *Immunological reviews* **219**: 37-54.

Rose, R.W., T. Bruser, J.C. Kissinger & M. Pohlschroder, (2002) Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol Microbiol* **45**: 943-950.

Rosenkrands, I., A. King, K. Weldingh, M. Moniatte, E. Moertz & P. Andersen, (2000) Towards the proteome of *Mycobacterium tuberculosis*. *Electrophoresis* **21**: 3740-3756.

Russell, D.G., C.E. Barry, 3rd & J.L. Flynn, (2010) Tuberculosis: what we don't know can, and does, hurt us. *Science* **328**: 852-856.

Saint-Joanis, B., C. Demangel, M. Jackson, P. Brodin, L. Marsollier, H. Boshoff & S.T. Cole, (2006) Inactivation of Rv2525c, a Substrate of the Twin Arginine Translocation (Tat) System of *Mycobacterium tuberculosis*, Increases {beta}-Lactam Susceptibility and Virulence. *J Bacteriol* **188**: 6669-6679.

Sakula, A., (1983) Robert koch: centenary of the discovery of the tubercle bacillus, 1882. *The Canadian veterinary journal. La revue veterinaire canadienne* **24**: 127-131.

Sander, P., M. Rezwan, B. Walker, S.K. Rampini, R.M. Kroppenstedt, S. Ehlers, C. Keller, J.R. Keeble, M. Hagemeier, M.J. Colston, B. Springer & E.C. Bottger, (2004) Lipoprotein processing is required for virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* **52**: 1543-1552.

Sani, M., E.N. Houben, J. Geurtsen, J. Pierson, K. de Punder, M. van Zon, B. Wever, S.R. Piersma, C.R. Jimenez, M. Daffe, B.J. Appelmelk, W. Bitter, N. van der Wel & P.J. Peters, (2010) Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathog* **6**: e1000794.

Sassetti, C.M. & E.J. Rubin, (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**: 12989-12994.

Schatz, A., E. Bugie & S.A. Waksman, (1944) Streptomycin, a substance exhibiting antibiotic activity against Gram-positive and Gram-negative bacteria. *Proc Exp Biol Med* **55**: 66–69.

Schmidt, F., S.S. Scharf, P. Hildebrandt, M. Burian, J. Bernhardt, V. Dhople, J. Kalinka, M. Gutjahr, E. Hammer & U. Volker, (2010) Time-resolved quantitative proteome profiling of host-pathogen interactions: the response of *Staphylococcus aureus* RN1HG to internalisation by human airway epithelial cells. *Proteomics* **10**: 2801-2811.

Schmidt, F. & U. Volker, (2011) Proteome analysis of host-pathogen interactions: Investigation of pathogen responses to the host cell environment. *Proteomics* **11**: 3203-3211.

Schnappinger, D., G.K. Schoolnik & S. Ehrt, (2006) Expression profiling of host pathogen interactions: how *Mycobacterium tuberculosis* and the macrophage adapt to one another. *Microbes Infect* **8**: 1132-1140.

Selengut, J.D., D.H. Haft, T. Davidsen, A. Ganapathy, M. Gwinn-Giglio, W.C. Nelson, A.R. Richter & O. White, (2007) TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res* **35**: D260-264.

Senaratne, R.H., B. Sidders, P. Sequeira, G. Saunders, K. Dunphy, O. Marjanovic, J.R. Reader, P. Lima, S. Chan, S. Kendall, J. McFadden & L.W. Riley, (2008) *Mycobacterium tuberculosis* strains disrupted in mce3 and mce4 operons are attenuated in mice. *Journal of medical microbiology* **57**: 164-170.

Shimono, N., L. Morici, N. Casali, S. Cantrell, B. Sidders, S. Ehrt & L.W. Riley, (2003) Hypervirulent mutant of *Mycobacterium tuberculosis* resulting from disruption of the mce1 operon. *Proc Natl Acad Sci U S A* **100**: 15918-15923.

Siegrist, M.S., B.M. Swarts, D.M. Fox, S.A. Lim & C.R. Bertozzi, (2015) Illumination of growth, division and secretion by metabolic labeling of the bacterial cell surface. *FEMS Microbiol Rev* **39**: 184-202.

Siegrist, M.S., M. Unnikrishnan, M.J. McConnell, M. Borowsky, T.Y. Cheng, N. Siddiqi, S.M. Fortune, D.B. Moody & E.J. Rubin, (2009) Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. *Proc Natl Acad Sci U S A* **106**: 18792-18797.

Solis, N. & S.J. Cordwell, (2011) Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. *Proteomics* **11**: 3169-3189.

Sonnhammer, E.L., G. von Heijne & A. Krogh, (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**: 175-182.

Sullivan, J.T., E.F. Young, J.R. McCann & M. Braunstein, (2012) The *Mycobacterium tuberculosis* SecA2 system subverts phagosome maturation to promote growth in macrophages. *Infect Immun* **80**: 996-1006.

Sutcliffe, I.C. & D.J. Harrington, (2004) Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiol Rev* **28**: 645-659.

Talaat, A.M., R. Lyons, S.T. Howard & S.A. Johnston, (2004) The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. *Proc Natl Acad Sci U S A* **101**: 4602-4607.

Tanrikulu, I.C., E. Schmitt, Y. Mechulam, W.A. Goddard, 3rd & D.A. Tirrell, (2009) Discovery of *Escherichia coli* methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc Natl Acad Sci U S A* **106**: 15285-15290.

Taylor, P.D., C.P. Toseland, T.K. Attwood & D.R. Flower, (2006) TATPred: a Bayesian method for the identification of twin arginine translocation pathway signal sequences. *Bioinformation* **1**: 184-187.

Taylor, R.K., V.L. Miller, D.B. Furlong & J.J. Mekalanos, (1987) Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc Natl Acad Sci U S A* **84**: 2833-2837.

Tiemersma, E.W., M.J. van der Werf, M.W. Borgdorff, B.G. Williams & N.J. Nagelkerke, (2011) Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PLoS One* **6**: e17601.

Twine, S.M., N.C. Mykytczuk, M.D. Petit, H. Shen, A. Sjostedt, J. Wayne Conlan & J.F. Kelly, (2006) In vivo proteomic analysis of the intracellular bacterial pathogen, *Francisella tularensis*, isolated from mouse spleen. *Biochem Biophys Res Commun* **345**: 1621-1633.

Uchida, Y., N. Casali, A. White, L. Morici, L.V. Kendall & L.W. Riley, (2007) Accelerated immunopathological response of mice infected with *Mycobacterium tuberculosis* disrupted in the mce1 operon negative transcriptional regulator. *Cell Microbiol* **9**: 1275-1283.

van der Woude, A.D., E.J. Stoop, M. Stiess, S. Wang, R. Ummels, G. van Stempvoort, S.R. Piersma, A. Cascioferro, C.R. Jimenez, E.N. Houben, J. Luirink, J. Pieters, A.M. van der Sar & W. Bitter, (2014) Analysis of SecA2-dependent substrates in *Mycobacterium marinum* identifies protein kinase G (PknG) as a virulence effector. *Cell Microbiol* **16**: 280-295.

von Heijne, G., (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol* **225**: 487-494.

Wang, M., C. Herrmann, M. Simonovic, D. Szklarczyk & C. von Mering, (2015) Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues and cell-lines. *Proteomics*.

Wang, M., M. Weiss, M. Simonovic, G. Haertinger, S.P. Schrimpf, M.O. Hengartner & C. von Mering, (2012) PaxDb, a database of protein abundance averages across all three domains of life. *Molecular & cellular proteomics : MCP* **11**: 492-500.

Wolfe, L.M., S.B. Mahaffey, N.A. Kruh & K.M. Dobos, (2010) Proteomic definition of the cell wall of *Mycobacterium tuberculosis*. *J Proteome Res* **9**: 5816-5826.

WorldHealthOrganization, (2014) Global Tuberculosis Report 2014.

Xia, Q., T. Wang, F. Taub, Y. Park, C.A. Capestany, R.J. Lamont & M. Hackett, (2007) Quantitative proteomics of intracellular *Porphyromonas gingivalis*. *Proteomics* **7**: 4323-4337.

Xiong, Y., M.J. Chalmers, F.P. Gao, T.A. Cross & A.G. Marshall, (2005) Identification of Mycobacterium tuberculosis H37Rv integral membrane proteins by one-dimensional gel electrophoresis and liquid chromatography electrospray ionization tandem mass spectrometry. *J Proteome Res* **4**: 855-861.

Yang, Y., M. Hu, K. Yu, X. Zeng & X. Liu, (2015) Mass spectrometry-based proteomic approaches to study pathogenic bacteria-host interactions. *Protein & cell*.

CHAPTER 2: PROBING FOR BACTERIAL PROTEINS AT THE HOST-PATHOGEN INTERFACE[1]

Introduction

Historically, studies of bacterial pathogens have largely relied on bacteria that are grown in the laboratory (i.e. under *in vitro* conditions); however, bacteria grown in the laboratory behave differently than they do in the host environment. Studying bacterial pathogens during infection has and will continue to help identify the virulence factors necessary for surviving in the host environment and establishing infection. One technological advance which proves the importance of studying pathogens directly in the host (i.e. under *in vivo* conditions) is IVET (*in vivo* expression technology). IVET uses a promoter trap approach to identify bacterial genes induced during infection. Fragments of genomic DNA are cloned into a plasmid, upstream of a reporter gene required for bacterial growth during infection (Mahan *et al.*, 1993; Angelichio & Camilli, 2002). Plasmids containing promoters active during infection drive expression of the reporter and allow for survival of those select bacterial clones *in vivo* (Mahan *et al.*, 1993). Clones that express active reporters during infection are then tested for reporter activity during *in vitro* growth, resulting in the successful identification of genes whose expression is specifically induced *in vivo*, many of which are virulence factors (Mahan *et al.*, 1993; Mahan *et al.*, 2000). Prominent examples of *in vivo* induced virulence factors include the *Salmonella* pathogenicity island 2 (SPI-2) encoded Type III secretion system, the *Legionella* ICM/Dot Type IV secretion

[1] Adapted for this dissertation from: Perkowski EF, Weerakoon D, Hayden JD, Ioerger TR, Oreper D, Gomez S, Sacchettini JC, Braunstein M. Probing for bacterial proteins at the host-pathogen interface. In preparation.

system, and the two component regulatory system PhoP in *Salmonella* (Mahan *et al.*, 2000). IVET was called a "bellwether", the first of new technologies to study pathogens directly in the host and improve our understanding of host-pathogen interactions (Barinaga, 1993). While decades have passed since the introduction of IVET and other *in vivo* technologies, critical questions remain about how pathogens establish disease, and new methods are required to further probe the host-pathogen interface *in vivo*.

Exported proteins are the proteins that are transported out of the bacterial cytoplasm to the bacterial cytoplasmic membrane, cell envelope, or out into the host environment. These proteins are located in an ideal place to interact with or protect from the host and to influence infection. Consequently, many exported proteins have functions in virulence (Isaac & Isberg, 2014; Agbor & McCormick, 2011; Ligon *et al.*, 2012). For this reason, decades of research has been devoted to identifying exported proteins of bacterial pathogens as candidates for being new virulence factors. However, all such studies have been limited by the available technologies, which study laboratory (*in vitro*) grown bacteria (de Souza & Wiker, 2011). In this Chapter, we describe the development of a novel method that we refer to as EXIT (EXported In vivo Technology) to specifically identify proteins exported by a bacterial pathogens during *in vivo* infection. Here, we used EXIT to successfully identify proteins exported by the human pathogen *Mycobacterium tuberculosis* during mouse infection. In the future, this technology could be applied to study diverse bacteria-host interactions.

Current Methods for Identifying Exported Proteins

Many bioinformatics programs are available to predict exported proteins. These programs search protein sequences for hallmarks of exported proteins (putative transmembrane helices and/or signal peptides) (Krogh *et al.*, 2001; Sonnhammer *et al.*, 1998; Hofmann & Stoffel, 1993;

von Heijne, 1992; Petersen *et al.*, 2011; Nakai & Horton, 1999; Hiller *et al.*, 2004; Taylor *et al.*, 2006; Juncker *et al.*, 2003; Sutcliffe & Russell, 1995; Rose *et al.*, 2002; Bendtsen *et al.*, 2005; Selengut *et al.*, 2007). However, these *in silico* methods are not fail-proof. There is significant disagreement between algorithms (data not shown; McDonough *et al.*, 2008) and the potential for incorrect assignment exists. In addition, these programs are limited to predicting conventional exported proteins and unable to identify exported substrates of specialized export systems. Thus, it is important to follow up *in silico* predictions of export with experimental validation.

There are two methods commonly used to experimentally discover new exported proteins. One method involves identification of proteins in exported subcellular fractions, often using mass spectrometry (MS). For this analysis, subcellular fractions containing exported proteins (membrane, cell wall, or secreted fractions) must first be prepared by physical (e.g. ultracentrifugation) or chemical (e.g. detergent) separation. MS based approaches to identify exported proteins are not currently applicable to bacterial pathogens during *in vivo* host infection (Schmidt & Volker, 2011). MS analysis is biased by abundance, which is particularly problematic when attempting to identify relatively rare bacterial proteins secreted into the host cell (Schmidt & Volker, 2011; Yang *et al.*, 2015; Kruh *et al.*, 2010). In fact, analysis of crude lysates from infected organs results in a very limited number of bacterial proteins identified (Kruh *et al.*, 2010). An alternate method is to first isolate/enrich bacteria from infected tissue, increasing the number of total bacterial proteins identified (Xia *et al.*, 2007; Twine *et al.*, 2006; Becker *et al.*, 2006; Schmidt *et al.*, 2010; Liu *et al.*, 2012; Pieper *et al.*, 2009). Isolation of bacterial cells from infected host cells could be useful for identifying proteins exported to the bacterial cell envelope. However, currently it cannot provide sufficient material for separation of exported fractions from cytoplasmic material (Schmidt & Volker, 2011; Yang *et al.*, 2015).

Therefore, application of MS based approaches to identification of *in vivo* exported proteins is not currently feasible.

An alternative strategy to biochemical identification of exported proteins is the use of genetic reporters (GR) of protein export. A protein export reporter is an enzyme whose activity or function depends on the reporter being exported/localized out of the cytoplasm. In the EXIT methodology developed in this study, we took advantage of the 'BlaTEM reporter. The *E. coli* β-lactamase BlaTEM cleaves β-lactam antibiotics and is able to confer β-lactam resistance on cells, as long as it gets exported out of the cytoplasm. The 'BlaTEM reporter is advantageous because it can be used as a selectable marker. Removal of the native signal peptide of BlaTEM ('BlaTEM) prevents export. However, when the truncated 'BlaTEM reporter is fused to a signal peptide, or the extra-cytoplasmic portion of an exported protein, export of the reporter confers β-lactam resistance (Figure 1.6). Thus, membrane, cell wall, or fully secreted proteins can be identified with the 'BlaTEM reporter (Figure 2.1). Only fusions to the periplasmic domains of membrane proteins will produce β-lactam resistance (Figure 2.1). This was first worked out in *Escherichia coli* (Broome-Smith & Spratt, 1986), and holds true when the 'BlaTEM reporter is used in the *ΔblaC* mutant of *M. tuberculosis* (McCann *et al.*, 2007), which lacks the endogenous exported β-lactamase BlaC of *Mycobacterium tuberculosis* (McCann *et al.*, 2007; McCann *et al.*, 2011). A transposon carrying the 'BlaTEM reporter was previously used to identify 111 *M. tuberculosis* proteins that are exported when the bacteria are growing on β-lactam containing agar plates *in vitro* (McCann *et al.*, 2011).

In general, MS based methods tend to have high false positive rates (high numbers of non-exported proteins identified as exported) and GR based methods tend to have high false

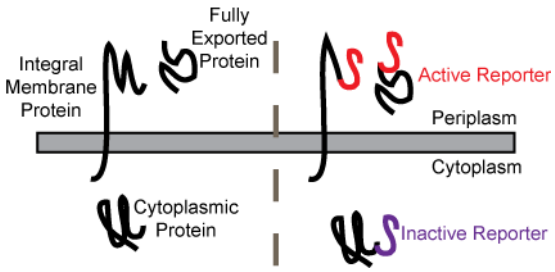

Figure 2.1.The ‘BlaTEM reporter. The ‘BlaTEM reporter is compatible with proteins localized to the bacterial cell membrane, periplasm, cell wall, or fully exported/secreted out of the bacterial cell. On the left are three example proteins, one cytoplasmic, one integral membrane, and one fully exported protein. The right panel demonstrates ‘BlaTEM fusions that will be active or inactive dependent on fusion location. Fusions located in frame in extra-cytoplasmic, or exported, portions of proteins will produce active fusions (red) and will confer β-lactam resistance to a cell producing this fusion. Fusions located within cytoplasmic proteins, or the cytoplasmic face of membrane proteins, will be inactive (purple) and will not confer β-lactam resistance.

negative rates (much smaller number of proteins identified, and thus many real exported proteins not identified). GR based approaches are highly accurate at identifying exported proteins, with >95% of proteins identified by GRs containing predicted export signals (Braunstein *et al.*, 2000; McDonough *et al.*, 2008; McCann *et al.*, 2011; Gomez *et al.*, 2000). However, GR based approaches historically identify far fewer proteins overall, in part because they are more labor intensive, and thus have a higher rate of false negative predictions than MS based approaches. MS based approaches, in comparison, tend to identify far more proteins but they have much higher false positive rates than GR based approaches. MS based approaches are less accurate than GRs at identifying proteins with predicted export signals. In *M. tuberculosis* MS based studies, it is often the case that 30% or less of identified proteins in an exported fraction contain a predicted export signal because the high level of sensitivity leads to identification of abundant cytoplasmic contaminants in exported fractions (Gunawardena *et al.*, 2013; Bell *et al.*, 2012; Malen *et al.*, 2011; de Souza & Wiker, 2011; Feltcher *et al.*, 2015). Neither MS or GR based approaches have been used to identify proteins exported during infection. As a result, the exported proteomes of pathogens during infection remain to be defined.

Protein Export in M. tuberculosis

M. tuberculosis exports proteins through two conserved and several specialized accessory systems. The Sec system and the Twin Arginine Translocation (Tat) pathway are conserved systems that export the bulk of proteins (Ligon *et al.*, 2012). Proteins exported by each of these systems are distinguished by the presence of N-terminal signal peptides which are cleaved off after export. Signal peptides directing proteins to the Tat system contain a RR motif (twin arginine) for which the system is named. Accessory systems (SecA2 and ESX1-5) export a smaller and not yet fully defined subset of proteins, and the precise requirements for export

through these pathways are only beginning to be discovered (Ligon *et al.*, 2012). SecA2 substrates fall into two groups, those that contain Sec-like N-terminal signal peptides and those that have no predicted N-terminal signal peptide (Gibbons *et al.*, 2007; Feltcher *et al.*, 2015; Braunstein *et al.*, 2003; van der Woude *et al.*, 2014). While the signal peptide is required for export of a SecA2 substrate if it exists, features of the mature domain of SecA2 exported proteins is what targets them for SecA2 dependent export (Feltcher *et al.*, 2013). ESX export requires a YxxxD/E motif in the exported substrate or a co-secreted partner protein, although elements outside of this motif define which ESX system will export the substrate (Daleke *et al.*, 2012). Overall, identifying the exported substrates of accessory systems has been challenging in *M. tuberculosis*, often due to our limited understanding of these systems.

The ‘BlaTEM reporter is compatible with export through both the Sec and Tat systems, and it is compatible with integral membrane proteins (McCann *et al.*, 2011; McCann *et al.*, 2007). The ‘BlaTEM reporter is also compatible with export through T3SS and T4SS (Ehsani *et al.*, 2009), and while *M. tuberculosis* is not predicted to encode either system it may possess T4SS-like proteins encoding a potential Tight adherence-like (Tad-like) secretion system (Danelishvili *et al.*, 2010; Tomich *et al.*, 2007). It remains unknown whether the ‘BlaTEM reporter is compatible with the accessory systems of export in *M. tuberculosis*.

Identification of Proteins Exported During Infection

We hypothesized that possibly the most important exported proteins are those that are only exported in the host environment. Therefore, we set out to develop a way to study protein export in the host and identify new exported proteins that had been long overlooked. Proteins exported only during infection could be regulated at the level of transcription, post-transcription, or export. Because β-lactams are already used to treat bacterial infections in mice, we

hypothesized that the ‘BlaTEM reporter could be used *in vivo* in a mammalian host. Consistent with this hypothesis, proof of principle studies demonstrated that ‘BlaTEM reporter fusions can be used with *M. tuberculosis* to identify exported proteins during infection of cultured macrophages (McCann *et al.*, 2007). We developed a selection strategy compatible with use of the ‘BlaTEM reporter during murine infection by optimizing β-lactam treatment for β-lactam sensitive *M. tuberculosis*, and analysis by next generation sequencing. In this study we applied this new method, EXported *In vivo* Technology (EXIT), to identify proteins exported by *M. tuberculosis* during murine infection.

EXIT identified 593 proteins as being exported by *M. tuberculosis* during infection, representing 57% of the *in silico* predicted exported proteome of *M. tuberculosis.* 100 of these proteins were experimentally demonstrated as being exported for the first time. Analysis of the exported fusions provided insight into the topology of several known virulence factors in *M. tuberculosis*. Further, several interesting classes of proteins were identified, including proteins with no predicted export signal and exported fusions in previously unannotated genomic regions. Additionally, 38 of the proteins identified by EXIT were exported significantly more during *in vivo* infection than *in vitro*. Because *M. tuberculosis* regulates production/export of these proteins in a spatial or temporal way, we hypothesize that they may be previously unidentified virulence factors.

Results

Our goal was to develop a method to identify proteins that were exported by *M. tuberculosis* during murine infection. The EXIT strategy was composed of four fundamental steps (Figure 2.2). In **Step 1**, a comprehensive library of $5x10^6$ *M. tuberculosis* clones was constructed carrying plasmids with random fragments of *M. tuberculosis* genomic DNA fused to

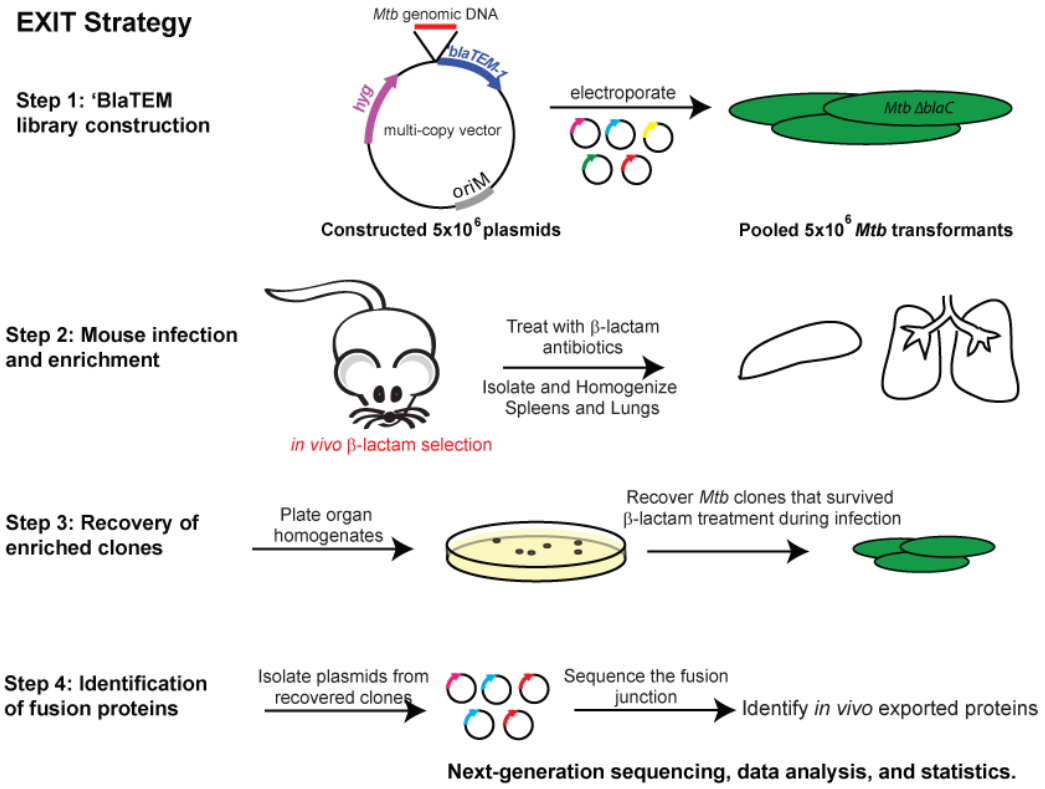

Figure 2.2 The EXported In vivo Technology (EXIT) Strategy. The EXIT strategy was composed of four fundamental steps. In **Step 1**, a comprehensive library *M. tuberculosis* clones carrying plasmids with random fragments of *M. tuberculosis* genomic DNA was fused to the *'blaTEM* reporter. First, a multi-copy plasmid was constructed containing origins of replication for mycobacteria and *E. coli*, a hygromycin resistance cassette, and a ClaI restriction enzyme cloning site upstream of the '*blaTEM* reporter gene. This plasmid contained no promoter upstream of the reporter, therefore an active reporter required in frame fusion to a protein that was actively expressed off its native promoter. Genomic DNA from a *ΔblaC* β-lactamase sensitive mutant of *M. tuberculosis* was fragmented by partial digestion four-base cutters AciI and HpaII. Ligation of digested fragments into the ClaI site of the plasmid was used to generate a library of $5x10^6$plasmids with unique DNA fusions. The plasmid library was transformed into the *ΔblaC* β-lactamase sensitive mutant of *M. tuberculosis* and $5x10^6$ transformants were pooled to generate the EXIT library. In **Step 2**, mice were infected with the EXIT library and treated with β-lactam antibiotics to select for EXIT clones exporting 'BlaTEM fusion proteins. Mice were infected by tail vein injection with approximately $4x10^6$ colony forming units (cfu). The remaining inoculum was collected and subjected to sequencing for the input as described below. Oral gavage treatment twice daily with amoxicillin and probenecid (β-lactam treatment) began one day after infection, and continued to 2 weeks after infection. Mice were sacrificed, and spleens and lungs were harvested and homogenized. In **Step 3**, organ homogenates were plated on solid agar media and grown for three weeks to recover *M. tuberculosis* clones that had survived β-lactam treatment during infection. Plates were scraped and colonies were pooled separately for lungs and spleens. In **Step 4**, plasmids from the recovered bacteria and the input samples were isolated and the fusion junction was sequenced using Illumina sequencing. Sequencing primers were designed to read out of the *'blaTEM* reporter and sequence the immediately adjacent fusion site. Sequences were trimmed and aligned to the *M. tuberculosis* genome corresponding to known AciI and HpaII restriction sites. Unique sequences were counted to identify the abundance of each fusion junction site within the population. The most highly abundant genes after *in vivo* β-lactam treatment were identified, which corresponded to plasmids producing 'BlaTEM reporters fused downstream of exported proteins.

the ‘BlaTEM reporter. In **Step 2**, mice were infected with the EXIT library and, starting one day after infection, treated by oral gavage with β-lactam antibiotics to select for clones exporting ‘BlaTEM fusion proteins *in vivo*. After 2 weeks of treatment, mice were sacrificed and spleens and lungs were harvested and homogenized. In **Step 3**, organ homogenates were plated on solid agar media and grown for three weeks to recover *M. tuberculosis* clones that had survived β-lactam treatment during infection. The recovered colonies were scraped from the agar plates and collected separately for lungs and spleens. In **Step 4**, library plasmids were isolated from the recovered bacteria surviving *in vivo* β-lactam treatment as well as from the input library and the fusion junctions were sequenced using Illumina sequencing. Sequencing primers were designed to read out of the ‘BlaTEM reporter and sequence the immediately adjacent *M. tuberculosis* fusion. Unique sequences were counted to identify the abundance of each fusion within the population. Finally, statistical modeling was used to identify highly abundant fusions recovered from the mice following *in vivo* β-lactam treatment.

***Construction of a comprehensive EXIT Library for* M. tuberculosis**

In order to construct the comprehensive plasmid library with random fragments of *M. tuberculosis* genomic DNA fused to the ‘BlaTEM reporter we first constructed plasmid pDW31 (Plasmid Table 2.1). Plasmid pDW31 is a multi-copy plasmid carrying an open reading frame (ORF) for the truncated ‘BlaTEM reporter along with a hygromycin resistance selectable marker, origins of replication for *Escherichia coli* and *M. tuberculosis*, and a ClaI restriction enzyme cloning site for cloning in random genomic fragments upstream of the reporter. There is no promoter sequence upstream of the reporter on plasmid pDW31; therefore, an active reporter requires an in frame fusion to a gene encoding an exported protein that is being expressed from its native promoter. Genomic DNA from a *ΔblaC* β-lactamase sensitive mutant of *M.*

tuberculosis was fragmented by partial digestion with 4 base pair cutters: AciI and HpaII restriction enzymes. Digestion was optimized for fragments between 500-5,000 base pairs in length, in an attempt to include native promoters for genes distally located in operons. Ligation of digested genomic DNA into the ClaI site of the plasmid followed by transformation into *E. coli* resulted in a plasmid library of 5 x10^6 unique plasmids. The plasmid library was subsequently purified from *E. coli* and then transformed into the *ΔblaC* β-lactam sensitive mutant of *M. tuberculosis*, resulting in a total of 5 x10^6 *M. tuberculosis* transformants (Strain Table 2.2).

We used a cumulative binomial equation to determine how many plasmids would be needed to have the whole *M. tuberculosis* genome represented in the EXIT library by at least one in frame fusion to the 'BlaTEM reporter. To achieve 99.9% confidence that any gene as small as 100 bp in length would be represented by at least one in frame fusion in the library required a library of 2 x10^6 unique plasmids (APPENDIX I). Therefore, the final EXIT library of 5 x10^6 plasmids theoretically had saturating coverage of the genome in frame with the reporter. The input EXIT library in *M. tuberculosis* was sequenced by next-generation sequencing using a primer at the fusion junction to the 'BlaTEM reporter (Primer Table 2.3). Sequence data confirmed saturation of the predicted AciI and HpaII sites within the genome. On average, the library contained a fusion every 26 base pairs in the *M. tuberculosis* genome, with the maximum non-represented region of the genome in the library being only 110 nucleotides long (Figure 2.3A).

Because out of frame and opposite orientation fusions are not able to produce a properly translated fusion of the 'BlaTEM reporter to a given Open Reading Frame (ORF), only one sixth of fusions are potentially informative. Despite this large reduction in usable fusions, the

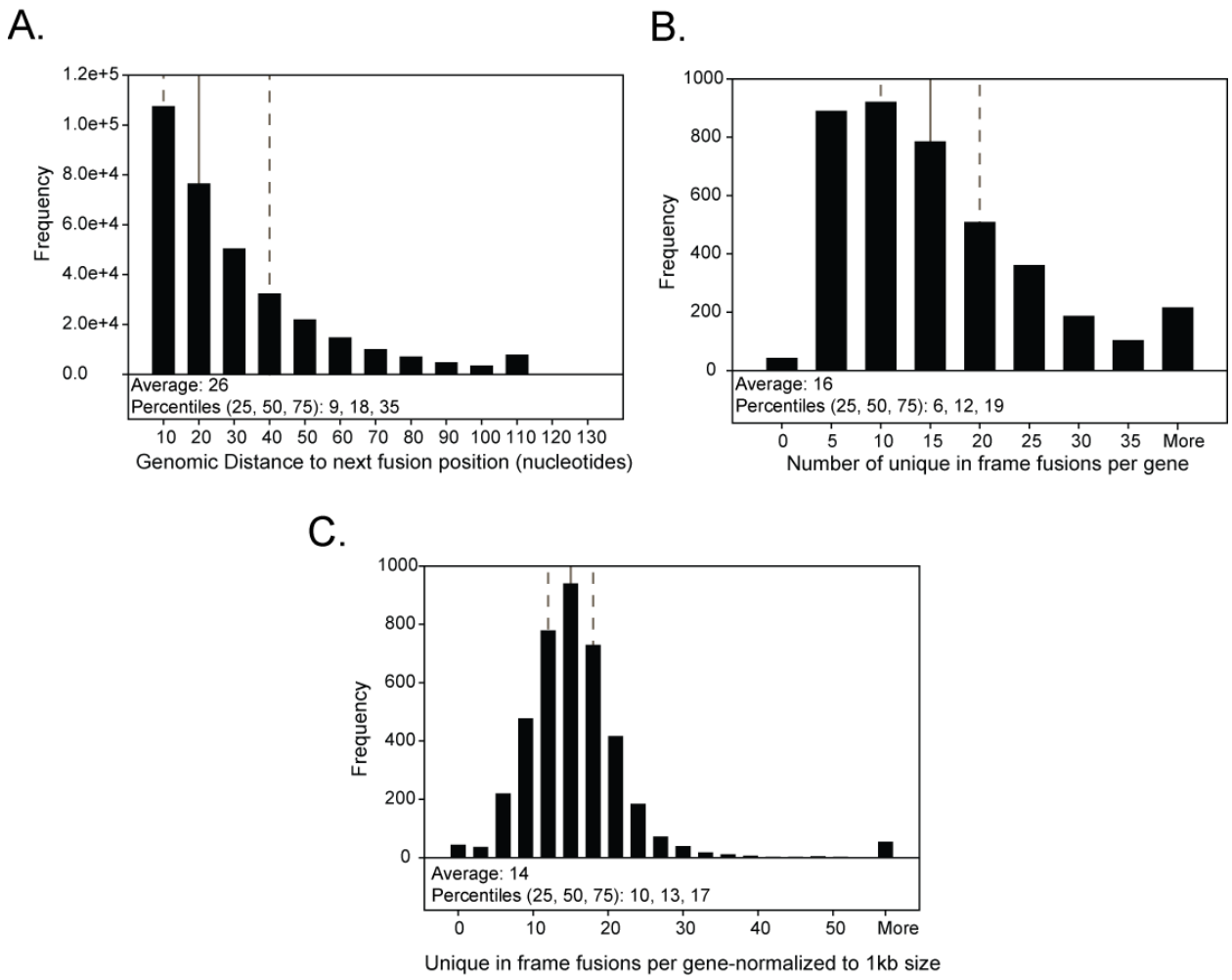

Figure 2.3 The EXIT library comprehensively represented the *Mycobacterium tuberculosis* proteome. A. The input library was sequenced by Illumina next-generation sequencing and analyzed to determine genome coverage of the EXIT library. Analysis of the frequency of fusion sites to the 'BlaTEM reporter determined that the largest gap of *M. tuberculosis* DNA between two unique fusion sites was 110 nucleotides. On average, there was a unique fusion site to the 'BlaTEM reporter every 26 nucleotides, demonstrating comprehensive coverage of the *M. tuberculosis* genome fused to the 'BlaTEM reporter. Percentiles are shown with dotted lines representing 25 and 75th percentile and a solid line representing the 50th percentile. **B.** Analysis of the frequency of in frame fusions identified an average of 16 in frame fusions per *M. tuberculosis* gene, with less than 1% of the genome not represented by an in frame fusion in the EXIT library. Percentiles are shown with dotted lines representing 25 and 75th percentile and a solid line representing the 50th percentile. **C.** The average size of annotated *M. tuberculosis* genes is 1006 nucleotides or approximately 1 kilobase (kb). The number of in frame fusions in each gene was normalized to a size of 1kb, and a histogram showed the frequency of in frame fusions per gene normalized to 1kb. The small population with greater than 50 in frame fusions per 1kb size was composed of GC rich proteins which contain disproportionally high numbers of AciI and HpaII restriction enzyme sites. Percentiles are shown with dotted lines representing 25 and 75th percentile and a solid line representing the 50th percentile.

complexity of the library was such that each gene was still represented by an average of 16 in-frame fusions, and some genes contained greater than 35 in-frame fusions (Figure 2.3B). Additionally, 99% of genes in the *M. tuberculosis* genome were represented by at least one in-frame fusion. The number of in-frame fusions per gene was directly proportional to gene length. When the length of each gene was normalized to 1kb, there were, on average, 14 in-frame fusions per 1 kb length (Figure 2.3C). Interestingly, there was a population of genes represented by greater than 50 in-frame fusions per kb in length. These genes were highly GC rich and thus contained a disproportionate number of AciI and HpaII restriction enzyme sites (Figure 2.3C).

Optimizing selection of exported 'BlaTEM-fusions in β-lactam treated mice

One of the major obstacles to developing EXIT was establishing the conditions for β-lactam selection in a murine host. *M. tuberculosis* is naturally β-lactam resistant due to an endogenous β-lactamase BlaC (Flores *et al.*, 2005). Because of this, EXIT was performed in a *M. tuberculosis ΔblaC* mutant background (Flores *et al.*, 2005). There were no established methods for β-lactam treatment of either wild type or *ΔblaC M. tuberculosis* during murine infection. Common strategies for treatment of mice with β-lactams include supplementing water with the antibiotics, single intra-muscular injections, and delivery of antibiotics by oral gavage. Although effective for the treatment of many different pathogens, and potentially useful for application of EXIT to other pathogens, supplementation of the water supply with amoxicillin or ampicillin was not sufficient to inhibit growth of β-lactam sensitive *M. tuberculosis* strains (data not shown). Likewise, treatment by oral gavage with amoxicillin alone was not sufficient (data not shown). Probenecid has long been used in conjunction with antibiotics to reduce drug efflux in the kidneys, increasing the serum concentration with few side effects (Robbins *et al.*, 2012). Addition of probenecid to an amoxicillin oral gavage treatment protocol (hereon referred to as β-

lactam treatment), and an increase in dosage frequency to twice daily oral gavage resulted in a successful treatment with evidence of killing the β-lactam sensitive *M. tuberculosis* strain.

To test the efficacy of our protocol we performed a proof of principle experiment to demonstrate β-lactam sensitivity of a strain exporting a functional reporter compared to a strain producing a non-exported reporter. The β-lactam sensitive *M. tuberculosis* strain produced a non-exported ‘BlaTEM reporter to control for any possible lysis and release of intracellular β-lactamase that could have potentially promoted survival. The β-lactam resistant *M. tuberculosis* strain produced the ‘BlaTEM reporter fused downstream of a signal peptide for an exported *M. tuberculosis* protein. Mice were infected with the β-lactam sensitive or β-lactam resistant strain, and bacterial burden in the spleen was determined after 14 days. Proof of principle experiments demonstrated a significant reduction in bacterial burden of a β-lactam sensitive *M. tuberculosis* strain in β-lactam treated mice (Figure 2.4). However, unlike the β-lactam sensitive strain, the β-lactam resistant strain grew normally despite β-lactam treatment (Figure 2.4). Importantly, both strains were fully virulent and grew to equal bacterial burden in untreated mice (Figure 2.4). These proof of principle experiments confirmed that the ‘BlaTEM reporter could be used in β-lactam treated mice to distinguish an exported protein from a non-exported protein.

Next we wanted to test if β-lactam resistant clones could be selectively enriched from complex mixture of β-lactam sensitive and β-lactam resistant clones, as would be the case with the EXIT library. In order to test this we infected mice with a mixture of β-lactam sensitive and β-lactam resistant *M. tuberculosis* (1% β-lactam resistant). After mice were infected with this 1% mixed population and treated with β-lactams for 14 days, 69% of the bacteria recovered from the spleen were β-lactam resistant (Table 2.4). This result was significant in demonstrating *in vivo* selection and enrichment of the β-lactam resistant fraction of the population while the β-lactam

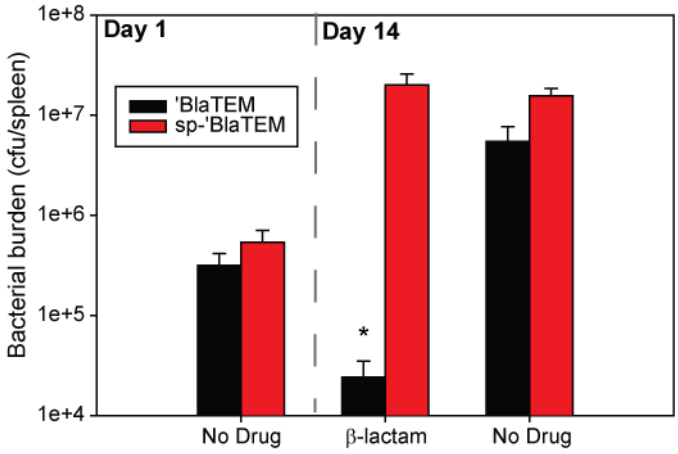

Figure 2.4. The ‘BlaTEM Reporter is Compatible with β-lactam Treatment During Murine Infection. Mice were infected by tail vein injection with *M. tuberculosis* strains producing a ‘BlaTEM reporter fused in frame with an exported signal peptide (sp-‘BlaTEM, red) or producing non-exported the ‘BlaTEM reporter alone (‘BlaTEM, black). One group of mice from each strain was sacrificed to determine initial bacterial burden in the spleens on day 1 after infection, and groups of mice were followed to day 14 after infection. Half of the mice were treated with the β-lactam antibiotic amoxicillin, and a synergistic drug probenecid twice daily by oral gavage, while half remained untreated. On day 14 the remaining mice were sacrificed and spleens were homogenized and plated on agar media to determine bacterial burden (colony forming units/cfu) with and without treatment. * represents statistical significance ($p<0.05$).

sensitive population was depleted. When mice were infected with the 1% mixed population but not treated with β-lactams, no enrichment for the β-lactam resistant population was observed, confirming that the enrichment observed in the treated animals was due to selection for β-lactam resistant bacteria (Table 2.4). Importantly, this confirmed that β-lactam resistant clones could be selectively enriched from a mixed population. In the input *M. tuberculosis* EXIT library 1% of the bacteria were β-lactam resistant *in vitro*. As with the above experiment, when mice were infected with the EXIT library and treated for 2 weeks with β-lactams, *in vivo* enrichment of β-lactam resistant clones was observed, going from 1% to 76% of the population, confirming our ability to select *in vivo* for β-lactam resistant clones.

Performing EXIT in M. tuberculosis infected mice

One day after intravenous infection, approximately 20% of the *M. tuberculosis* inoculum was found in the spleen and 1% in the lungs (data not shown), consistent with previous studies (Orme & Gonzalez-Juarrero, 2007). Prior to conducting large scale experiments, we calculated the number of animals that would be required in order to comprehensively screen the EXIT library in mice. Using a modified binomial calculation (APPENDIX I) we determined that infection of 24 mice with $5x10^6$ bacteria from the EXIT library would result in a 99.5% probability of any individual clone in the library establishing infection in the spleen of at least one mouse. With this knowledge, large scale EXIT experiments were conducted with 24 mice and were performed in duplicate on separate occasions. For each experiment, mice were intravenously infected with $5x10^6$ bacteria from the EXIT library and treatment with β-lactams started one day after infection. After two weeks, lungs and spleens were collected and organ homogenates were plated onto agar plates. Colonies for each organ were pooled separately, plasmids were isolated, and the fusion junctions were sequenced by Illumina next-generation

sequencing technology with a specific primer reading out of the ‘BlaTEM fusion junction. In parallel, the *M. tuberculosis* input library (inoculum) used to infect the mice was subjected to next-generation sequencing to determine the representation of individual fusions at the start of each experiment.

A unique pipeline was built for analysis of the sequencing data, which consisted of 136 million paired-end sequenced reads (see Methods). Briefly, sequences were trimmed and fusion junctions between the genomic DNA and the reporter were mapped to the genome. Unique reads for each fusion site were counted and the abundance was determined individually for each organ and experiment. Lastly, the gene fused immediately upstream of the reporter was determined to be in frame with the reporter, out of frame with the reporter, or outside of an annotated gene. Results from the two independent experiments were highly correlated as determined by a Pearson’s Product Moment Correlation, input: r=0.818, output from treated mice: r=0.784 (Figure 2.5A and 2.5B).

The majority of individual fusions were less abundant when rescued from treated mice (i.e. following *in vivo* β-lactam treatment) than they were in the input library (Figure 2.6A). This depletion was expected as the library includes many out of frame fusions or fusions with cytoplasmic proteins. A smaller population of sequenced fusions was enriched (Figure 2.6A). An early sign that EXIT was working as expected came when we directly compared the enrichment of in frame fusions with the reporter to out of frame fusions. Out of frame fusions should not produce β-lactamase activity. Therefore, out of frame fusions should not support growth in the presence in β-lactam treated mice and they should not be enriched. As expected, out of frame fusions were highly depleted during *in vivo* β-lactam selection, and 95% of the highly enriched fusions were in frame with the reporter.

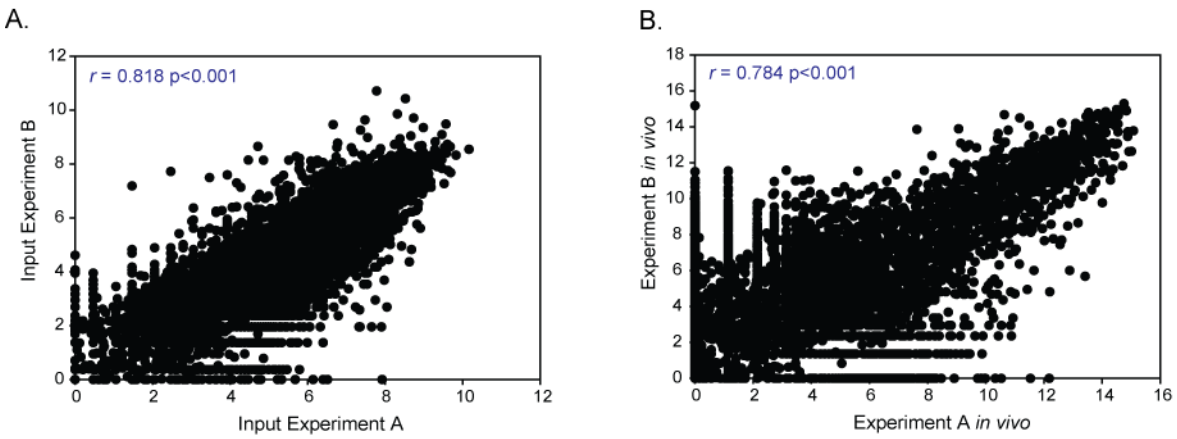

Figure 2.5. EXIT replicate experiments demonstrated high reproducibility and correlation. A. Raw count values in each input replicate experiment (A and B) were plotted for each fusion junction site on a log2 scale. A Pearson Product Moment correlation identified a significant correlation *r* value of 0.818. **B.** Raw count values in the mouse spleen output for each fusion junction site were plotted on a log2 scale for replicate experiments A and B. A Pearson Product Moment correlation identified a significant correlation *r* value of 0.784.

M. tuberculosis clones that expressed in frame fusions to exported proteins were expected to survive and replicate during *in vivo* β-lactam treatment, and be the most abundant clones in the output from treated mice. Abundance values were assigned for each gene based on the number of sequenced reads identified in the output from β-lactam treated mice. To impose the highest level of stringency in identifying genes that were reliably more abundant, the lowest abundance value for a gene from the two experimental replicates was selected for the subsequent statistical analysis (for details see Methods). A Gaussian mixture model was constructed to describe the two populations: a large group of low abundance genes and a smaller group of high abundance genes (Figure 2.6B). The statistical model set a cutoff of log10 of 2.90, or an abundance of approximately 800 unique sequenced reads, above which a given gene was identified in the high abundance group. Using this statistical analysis, 593 genes were identified as highly abundant after *in vivo* β-lactam treatment, and thus predicted to encode exported proteins (Figure 2.6B, APPENDIX II).

EXIT successfully identified exported proteins

As a first assessment of the accuracy of EXIT, we searched for export signals (signal peptides and transmembrane domains) in the 593 proteins identified (APPENDIX II). To facilitate identification of export signals in proteins identified in EXIT, we first analyzed the whole *M. tuberculosis* H37Rv proteome for transmembrane domains using TMHMM (Krogh *et al.*, 2001), signal peptides using Signal P (Petersen *et al.*, 2011), and lipoproteins and Tat signal peptides (Sutcliffe & Harrington, 2004; McDonough *et al.*, 2008) (APPENDIX III). There are many transmembrane and signal peptide prediction algorithms available. Our decision to the TMHMM program is based on the report of it having a lower frequency of false-positive predictions than other transmembrane prediction programs (Punta *et al.*, 2007). We chose the

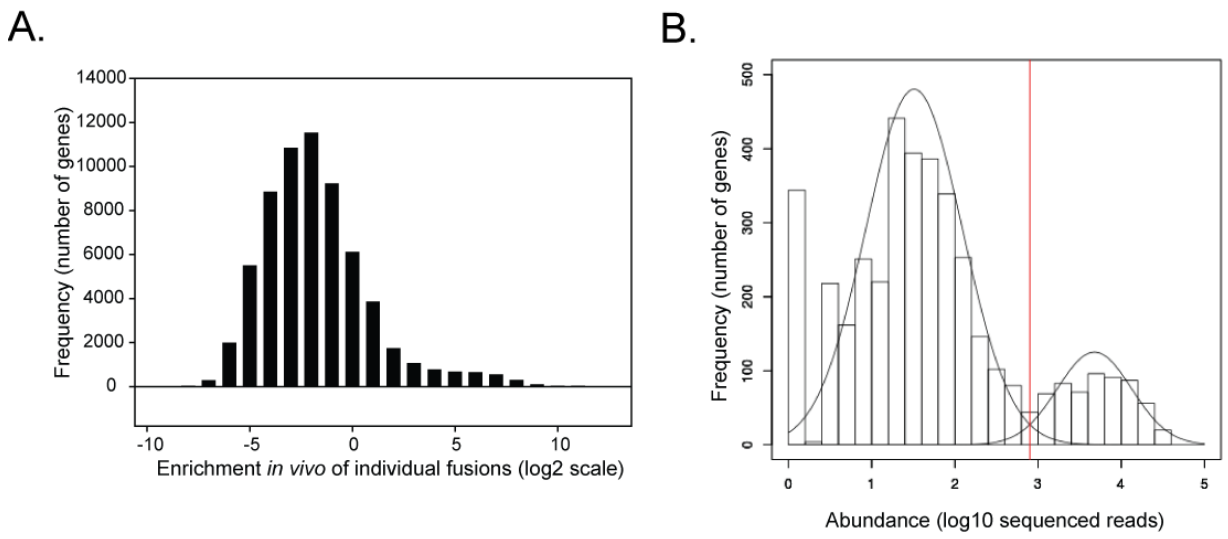

Figure 2.6. EXIT identified 593 proteins as exported during murine infection. A. The average abundance of each fusion site in the output from mice was divided by the average abundance in the input. Fusion sites with values of zero, occurring mostly in the output, were assigned a value of 1. A histogram was generated to compare the number of fusion sites enriched or depleted during *in vivo* β-lactam treatment. **B.** The most abundant fusion position within each annotated gene in the *M. tuberculosis* genome (3999 protein coding genes from H37Rv RefSeq genome annotation released January 9 2012) was identified individually within the output for each of two replicate experiments. The lower of these two numbers was plotted on a histogram. A two component Gaussian mixture model (black line overlay) was used to generate a statistical model distinguishing between high abundance genes (right) and low abundance genes (left), with a statistical cutoff of log10=2.90, or approximately 800 counts (red line). 593 genes were identified in the high abundance population corresponding to EXIT exported proteins.

Signal P program because it performs better than other tools in identifying signal peptides in *M. tuberculosis* exported proteins (Leversen *et al.*, 2009). Using these *in silico* programs to predict signal peptides and transmembrane domains, 26% (1040 proteins) of the *M. tuberculosis* proteome was predicted to be exported (APPENDIX III). This compares well to analyses across bacterial species, which usually predict 20-30% of the proteome to be exported (Saleh *et al.*, 2001).

Only 26% of proteins containing in frame fusions in the input library contained *in silico* predicted export signals. However, nearly all (95%) of the 593 proteins identified by EXIT as exported in the mouse were *in silico* predicted exported proteins (Figure 2.7A, APPENDIX II). This demonstrated strong selection *in vivo* for fusion proteins possessing export signals. Overall, 57% of the *in silico* predicted exported proteome of *M. tuberculosis* was identified as exported *in vivo* by EXIT. Proteins identified as exported *in vivo* included those with *in silico* predicted Sec signal peptides, Tat signal peptides, lipoproteins, and proteins with transmembrane domains, confirming broad compatibility with the ‘BlaTEM reporter (APPENDIX II).

As an additional test of the ability of EXIT to correctly identify exported proteins, we compared the list of 593 to lists of known exported (Table 2.5) and known non-exported proteins (Table 2.6). Because we had no information about *in vivo* exported proteins, we relied on a list of proteins known to be exported *in vitro*, presuming that many or all would also be exported *in vivo* (Table 2.5). Therefore we compared EXIT identified proteins to proteins experimentally demonstrated as exported *in vitro*, including well-known *M. tuberculosis* exported antigenic proteins like Ag85B and 19 kDa/LpqH (McCann *et al.*, 2011). 82% of these experimentally demonstrated exported proteins were identified as exported *in vivo* by EXIT. We also searched the list of 593 proteins for examples of known non-exported (cytoplasmic) proteins. Highly

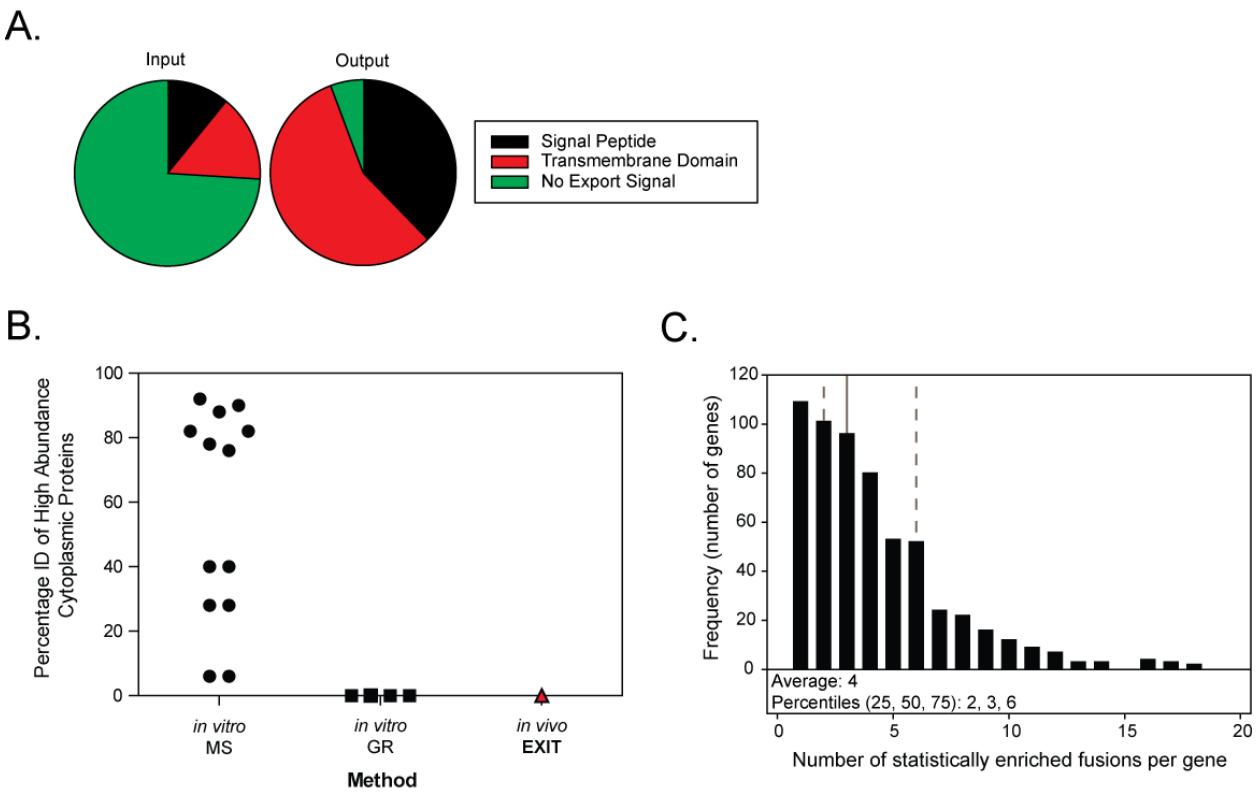

Figure 2.7. EXIT successfully and reliably identified exported proteins. **A.** The input EXIT library was composed of fusions into 99% of *M. tuberculosis* genes, with 74% encoding proteins with no predicted export signal (green), 15% encoding predicted integral membrane proteins (red), and 11% encoding proteins containing predicted signal peptides (black). In contrast, 95% of proteins in the EXIT output contained an export signal. The 593 proteins identified as exported in EXIT were composed of 57 % predicted integral membrane proteins (red), 38% of proteins containing a predicted signal peptide (black), and 5% of proteins with no predicted export signal (green).**B.** The top 50 most abundant cytoplasmic proteins were identified by PaxDB, a database calculating the relative abundance of proteins from published proteomics datasets (Wang *et al.*, 2015; Wang *et al.*, 2012). Proteins identified as abundant in PaxDB but containing predicted export signals (transmembrane domains and signal peptides were predicted by SignalP, TatP, TMHMM, (Sutcliffe & Harrington, 2004), and (McDonough *et al.*, 2008)), or proteins known to be exported (e.g. EsxA and EsxB) were excluded, to generate a list of the top 50 most abundant cytoplasmic proteins in *M. tuberculosis*. Exported proteins of *M. tuberculosis* have been identified on *in vitro* grown bacteria by mass-spectrometry (MS) based methods as well as use of genetic reporters (GR). MS based methods have analyzed proteins in exported fractions including the bacterial cell membrane, cell wall, and extracellular or culture filtrate fractions. For comparison, all lists were analyzed using the H37Rv RefSeq genome annotation released January 9 2012. The number of high abundance cytoplasmic proteins was determined for each MS and GR based study of exported proteins. **C.** Genes identified as encoding exported proteins were analyzed for the number of statistically enriched fusions after *in vivo* β-lactam treatment. On average, 4 unique fusion sites were enriched for each exported protein. Percentiles are shown with dotted lines representing 25 and 75th percentile and a solid line representing the 50th percentile.

abundant cytoplasmic proteins would be the most likely candidates for false positive identification by EXIT, as highly expressed intracellular BlaTEM fusions could potentially result in release of β-lactamase through a non-specific process. Using PaxDB, a database that mines *in vitro* proteomics datasets and estimates cellular abundance of proteins (Wang *et al.*, 2015; Wang *et al.*, 2012), we compiled a list of 50 of the most abundant cytoplasmic proteins in *M. tuberculosis* (Table 2.6). Not one of these 50 highly expressed cytoplasmic proteins were on the list of 593 proteins identified by EXIT (Figure 2.7B), thus we were very pleased at the apparent success of EXIT to identify a large number of exported proteins with a very low false positive rate. In comparison, MS based methodologies to identify proteins in exported subcellular fractions consistently identify more than 40 of the 50 most abundant cytoplasmic proteins identified by PaxDB (Figure 2.7B) (Gu *et al.*, 2003; Rosenkrands *et al.*, 2000a; Mawuenyega *et al.*, 2005; Wolfe *et al.*, 2010; Xiong *et al.*, 2005; Malen *et al.*, 2007; Malen *et al.*, 2011; Bell *et al.*, 2012; Gunawardena *et al.*, 2013).

Finally, we wanted to compare different fusion sites within EXIT identified proteins, because identification of multiple unique fusion sites as independently enriched would provide extra confidence that EXIT reliably identified the same proteins as exported. We looked at cases of multiple unique ‘BlaTEM fusion sites in the same gene and compared the abundance of each fusion in the input and output samples to determine statistical enrichment. 2,516 unique fusion sites were identified as statistically enriched during *in vivo* β-lactam treatment, corresponding to the 593 proteins identified as exported. Each protein was represented by an average of four unique enriched fusion sites (Figure 2.7C). This provided additional confidence in the power of EXIT to identify exported proteins, because 82% of the 593 proteins were identified by multiple unique fusion sites within each experiment.

EXIT identified new exported proteins

Proteins never previously experimentally shown as exported

Many EXIT identified proteins had previously been identified as exported *in vitro* by MS and GR based methods (Figure 2.8A), further validating the EXIT method for identification of exported proteins. However, despite multiple previous *in vitro* studies using MS and GR based methods (Gu *et al.*, 2003; Rosenkrands *et al.*, 2000a; Mawuenyega *et al.*, 2005; Wolfe *et al.*, 2010; Xiong *et al.*, 2005; Malen *et al.*, 2007; Malen *et al.*, 2011; Bell *et al.*, 2012; Gunawardena *et al.*, 2013; Braunstein *et al.*, 2000; McDonough *et al.*, 2008; McCann *et al.*, 2011; Gomez *et al.*, 2000), EXIT still identified 100 proteins that had no previous experimental evidence of export (Figure 2.8A).

Proteins not predicted *in silico* as exported

Due to the very high percentage (95%) of proteins identified by EXIT that possess *in silico* predicted export signals and the evidence of multiple unique *in vivo* enriched fusions per EXIT identified protein (Figure 2.7A, Figure 2.7C), proteins that were identified as exported by EXIT with no conventional export signal are compelling candidates for non-conventional exported proteins. EXIT identified 32 proteins (5% of EXIT predicted proteins) which lack a predicted export signal, and could be true exported proteins overlooked by current bioinformatics programs (Table 2.7). To test whether these 32 proteins include examples of true exported proteins, three representative proteins were engineered with an HA tag at the C-terminus (Rv1728c, Rv3707c, and Rv3811) and these genes were expressed off a constitutive *hsp60* promoter in *M. tuberculosis.* Cells were lysed by French press to generate whole cell lysates (WCL), and subcellular fractions containing the cell wall (CW), membrane (MEM) and soluble/cytoplasmic (SOL) proteins were generated by differential ultracentrifugation. Western

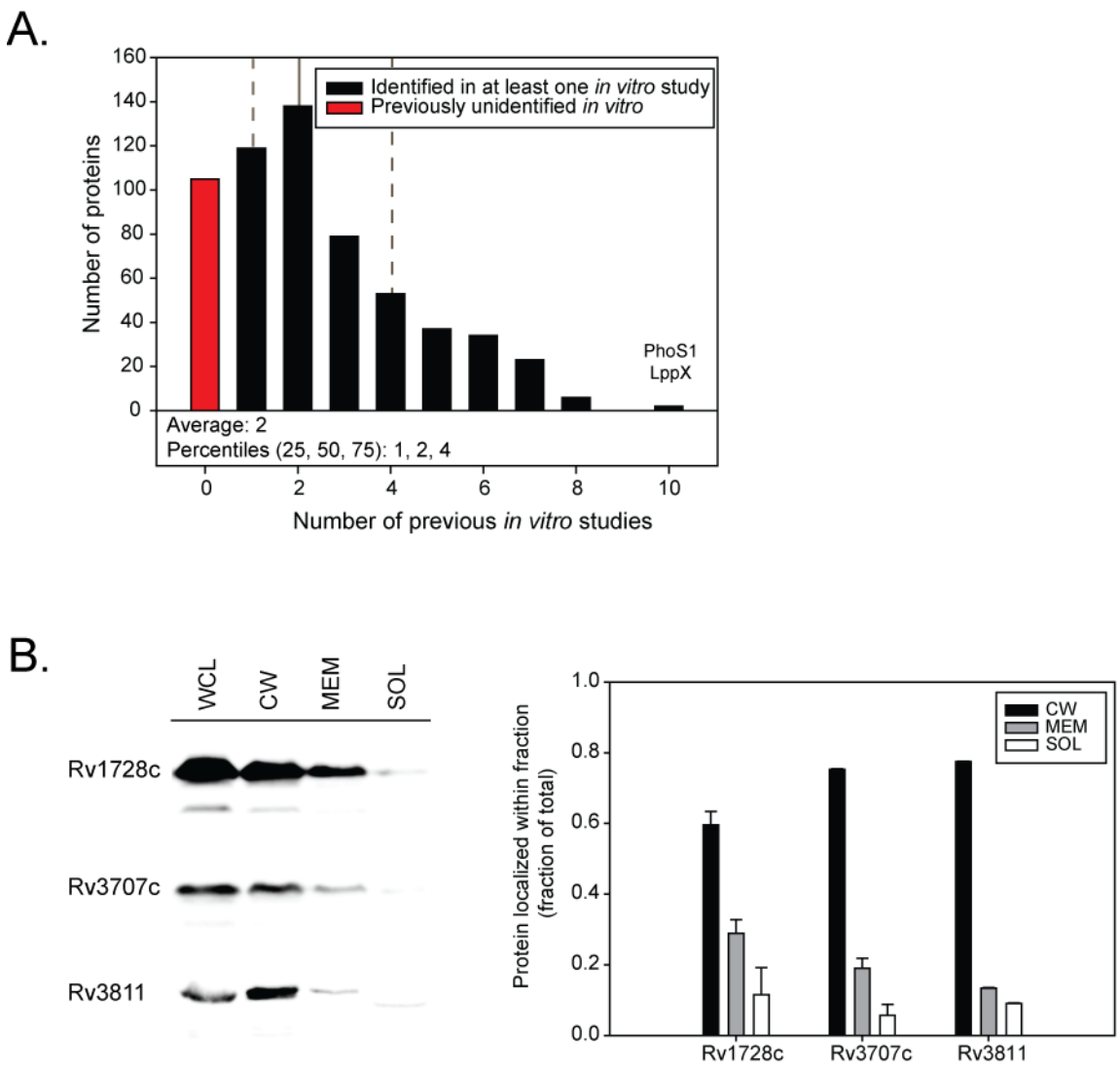

Figure 2.8. EXIT identified new exported proteins. A. Proteins identified as exported by EXIT were compared with previous studies to determine how many previous studies had identified them as exported. 100 proteins designated in red had not previously been identified as exported *in vitro*. Two proteins, LppX and PhoS1 were each identified in 10 studies. Percentiles are shown with dotted lines representing 25 and 75th percentile and a solid line representing the 50th percentile. **B.** Three proteins with no *in silico* predicted export signal (Rv1728c, Rv3707c, and Rv3811) were engineered with C-terminal HA tags, and produced in *M. tuberculosis*. Cells were irradiated, lysed by French pressure cell into whole cell lysate (WCL), equalized by BCA protein quantification, and fractionated by differential ultracentrifugation into cell wall (CW), membrane (MEM), and soluble/cytoplasmic (SOL) fractions. Fractions containing equal cellular material were separated by SDS-PAGE and HA tagged proteins were detected by Western blot with anti-HA antibodies. Chemiluminescent signal was quantified individually for each subcellular fraction by the ChemiDoc MP (Biorad) with Image Lab software (Biorad) and reported as fraction of total from total signal (SOL+CW+MEM). Error bars represent standard deviation from duplicate biological replicates.

blot analysis of the subcellular fractions demonstrated that all three of these proteins are exported to the cell wall fraction of *M. tuberculosis* (Figure 2.8B). These results speak to the ability of EXIT to identify exported proteins that are missed by the heavily relied upon *in silico* prediction tools for export signals. Nine of the proteins that EXIT identified as exported that lack predicted export signals have not previously been identified as exported by any available *in silico* or *in vitro* method (Table 2.7). This list of nine proteins included Rv3811, one protein we successfully validated as exported.

Potential exported ORFs in unannotated regions

The vast majority of enriched fusions were in frame with annotated genes. However, EXIT also identified enriched fusions in a small subset of unannotated regions of the genome, spanning intergenic spaces, suggesting that these regions may contain unannotated/misannotated exported proteins. One such enriched exported fusion was immediately upstream of the gene *rv3035*, which encodes one of the 32 EXIT identified proteins with no predicted export signal. Interestingly, upstream of the predicted start site for *rv3035* is an in-frame translational start site that would introduce a predicted N-terminal transmembrane domain and account for this additional fusion being identified by EXIT. This upstream sequence also provided a likely explanation for the export of *rv3035*.

Unannotated regions which contained multiple enriched fusions in the same reading frame and within 100 base pairs of each other were identified as the most likely to represent currently unannotated ORFs encoding exported proteins (Table 2.8). One genomic region downstream of Rv2307A contained six highly enriched fusion sites that were all in the same reading frame. We used the Geneious software program (Kearse *et al.*, 2012) to identify potential ORFs within this region, and a single ORF was consistent with all six enriched fusion sites.

Furthermore, there was an *in silico* predicted signal peptide at the N-terminus of the translated ORF. Future studies are required to better characterize these ORFs as new exported proteins. The identification of exported proteins lacking *in silico* predicted export signals and identification of unannotated exported proteins demonstrates the power of an unbiased genome-wide strategy as well as an unbiased sequencing platform.

EXIT fusions identify proteins exported in the lungs

Because of low (1%) seeding of the lungs following intravenous infection and the associated bottleneck issues, we were unable to comprehensively test the EXIT library in the lungs of mice. This was evident in data analysis of the fusions present in bacteria recovered from lungs of treated mice, where there was not a sufficient number of high abundance fusions identified to develop statistical models. However, using a threshold of 3.5 fold enrichment of a gene in the lungs over abundance in the input in duplicate experiments there were 282 proteins, representing the strongest candidates for being exported in the lungs. Of these, 274 (97%) were identified by EXIT as exported in the spleen (APPENDIX II). These 282 proteins represent 46% of the 593 proteins identified as exported in the spleen (APPENDIX II). We predict that a higher proportion of the 593 proteins were actually exported in the lungs, but they were not identified due to bottleneck effects.

There was an interesting group of eight proteins not identified by EXIT as exported in the spleen but predicted to be exported in the lungs (Table 2.9). Interestingly, 4 of these 8 proteins are PE_PGRS proteins, a poorly understood class of proline and glutamic acid repeat containing proteins unique to mycobacteria. The identification of these 4 PE_PGRS proteins in the lung data from EXIT but not in the more comprehensive spleen data is intriguing as it could reflect lung specific mechanisms of expression or export for these proteins. PE_PGRS proteins are a

subfamily of PE/PPE proteins (proteins containing Pro-Glu or Pro-Pro-Glu repeat domains) that additionally contain polymorphic guanosine-cytosine-rich sequences (PGRS) (Goldberg *et al.*, 2014). While the function of PE/PPE proteins remains uncertain, evidence suggests that they are surface localized, and members of the PE/PPE protein family have been shown to be important for *M. tuberculosis* virulence (Banu *et al.*, 2002; Brennan *et al.*, 2001; Forrellad *et al.*, 2013b). Identification of four PE_PGRS proteins as exported only in the lungs identifies them as interesting candidate virulence factors

EXIT exported fusions provide topology information for membrane proteins

The specific insertion sites of the ‘BlaTEM fusion identified by EXIT provides its own useful information about exported protein domains located outside of the cytoplasm. Such information is particularly informative for understanding functional domains of integral membrane proteins. Only fusions located within periplasmic domains of membrane proteins can provide β-lactam resistance; therefore, enriched exported fusion sites can be used to determine exported domains and the topology of membrane proteins. Several bioinformatics algorithms exist to predict which domains of transmembrane proteins will be exported to the periplasm (Krogh *et al.*, 2001; Hofmann & Stoffel, 1993; von Heijne, 1992; Claros & von Heijne, 1994; Jones *et al.*, 1994). These algorithms are guided by the positive-inside rule (von Heijne & Gavel, 1988); however, in terms of topology predictions for membrane proteins it is extremely common for different programs to disagree.

Because the EXIT library had an average of 16 in-frame fusions per gene, it provided information on the exported domains of integral membrane proteins in unprecedented detail. All 2,516 exported fusions sites for the 593 identified EXIT fusions are reported (APPENDIX II). There were several examples, discussed below, where the fusion sites provided key insight into

the topology of *M. tuberculosis* membrane proteins that are currently under extensive investigation for the important role(s) they play in virulence, physiology or drug resistance.

MmpL3 and MmpL family members

The MmpL (mycobacterial membrane protein Large) family of proteins is composed of polytopic integral membrane proteins, and *M. tuberculosis* H37Rv encodes 13 *mmpL* genes (Tekaia *et al.*, 1999). Many MmpL proteins contribute to *M. tuberculosis* virulence (Domenech *et al.*, 2005; Domenech *et al.*, 2004; Converse *et al.*, 2003) and MmpL3 is essential in mycobacteria (Domenech *et al.*, 2005). The substrates of most MmpL transporters have yet to be identified; however, all the MmpL transporters characterized to date act to export lipids (Domenech *et al.*, 2004; Varela *et al.*, 2012; Rayasam, 2014; Converse *et al.*, 2003). The essential MmpL3 exports mycolic acid precursors required for biosynthesis of the mycobacterial outer membrane (Rayasam, 2014). Resistance mutations to several new TB inhibitors map to the gene encoding MmpL3 (Li *et al.*, 2014; Rayasam, 2014; Remuinan *et al.*, 2013; Poce *et al.*, 2013; Tahlan *et al.*, 2012; La Rosa *et al.*, 2012), which has attracted a great deal of attention for MmpL3 being a potential novel drug target. MmpL proteins usually possess 11-12 predicted transmembrane domains with two large loops and a large C-terminal domain. However, the membrane topology of MmpL3, and other MmpL members, is under debate (Li *et al.*, 2014; Rayasam, 2014; Remuinan *et al.*, 2013; Poce *et al.*, 2013; Tahlan *et al.*, 2012; La Rosa *et al.*, 2012; Sandhu & Akhter, 2015; Varela *et al.*, 2012). Prediction programs disagree about the location (cytoplasmic or periplasmic) of the two large loops of the protein (Figure 2.9A), and multiple different topology predictions for this protein are published (Li *et al.*, 2014; Rayasam, 2014; Remuinan *et al.*, 2013; Poce *et al.*, 2013; Tahlan *et al.*, 2012; La Rosa *et al.*, 2012; Sandhu & Akhter, 2015; Varela *et al.*, 2012). EXIT identified 13 different exported fusions in the two

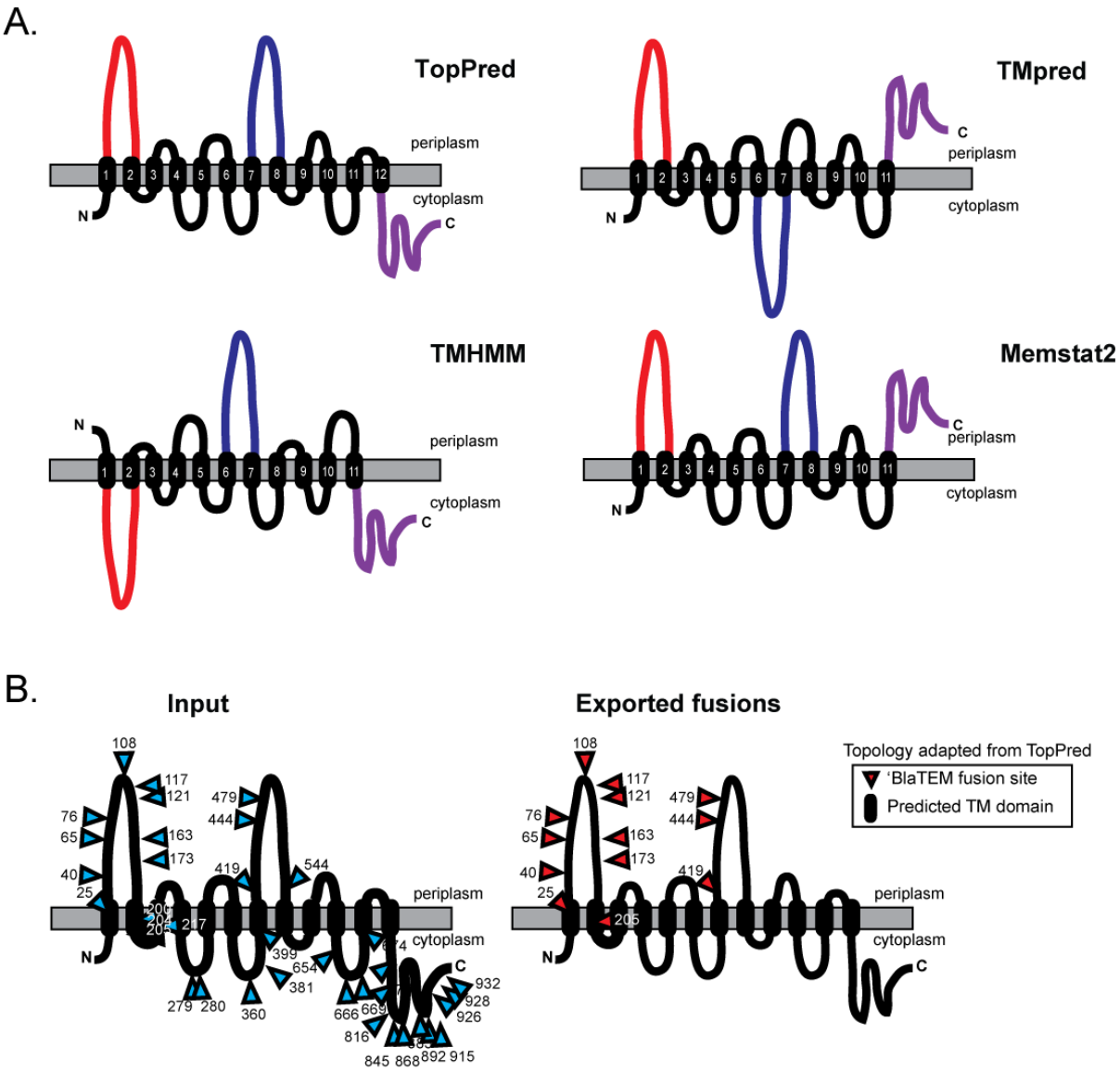

Figure 2.9. EXIT Exported fusions clarify the topology of membrane protein MmpL3. A. MmpL3 predictions for transmembrane domains prediction programs including TopPred, TMHMM, TMpred, and Memstat2 predict two large loops (red and blue) and a C-terminal domain (purple); however, disagree on the cellular location of these domains. **B.** Topology diagrams were created by choosing transmembrane predictions that best agreed with the exported fusion sites identified by EXIT. 37 unique fusions sites in MmpL3 were represented in the input library (blue triangles). Of these, 13 fusion sites were identified as exported in EXIT (red triangles top panel), corresponding to exported domains of the MmpL3 protein.

predicted large loops of MmpL3 (Figure 2.9B). These fusion sites serve to define the two large loops as being localized to the periplasm, which supports one of the topology predictions for MmpL3, generated by TopPred (Figure 2.9B) (Claros & von Heijne, 1994). We similarly analyzed the enriched fusions identified by EXIT for topological information for the remaining MmpL family members. Eleven MmpL family members were represented by 52 unique exported fusion sites within the large loops of the proteins, again indicating that MmpL family proteins have the two large loops positioned on the periplasmic side of the cytoplasmic membrane (Figure 2.10). This topology is consistent with the expected topology for RND (Resistance, Nodulation, Division) family transporters, with which MmpL proteins share some homology (Domenech *et al.*, 2005). The composite topology for MmpL proteins (Figure 2.10) did not include data for MmpL6, which was not identified as exported in EXIT. Additionally, the composite topology did not include MmpL10, because EXIT identified exported fusion sites contradicted all *in silico* predicted topologies based on *in silico* predicted transmembrane domains. MmpL10 may be structured differently than other MmpL proteins; however, further studies are necessary to determine the location of transmembrane domains in MmpL10 that may explain the exported fusion sites identified by EXIT.

MmpL proteins are often encoded by genes in operons containing *mmpS* genes (mycobacterial membrane protein small) (Tekaia *et al.*, 1999). Several MmpS proteins have recently been shown to function in MmpL mediated transport (Wells *et al.*, 2013; Jones *et al.*, 2014). Most MmpS proteins are predicted to contain a single N-terminal transmembrane domain (Krogh *et al.*, 2001). Analysis of unique MmpS fusions identified a consensus topology where the C-terminal domain was periplasmic (Figure 2.10). MmpS1 is predicted to have a second C-terminal transmembrane domain, and an exported fusion localizes the loop between the two

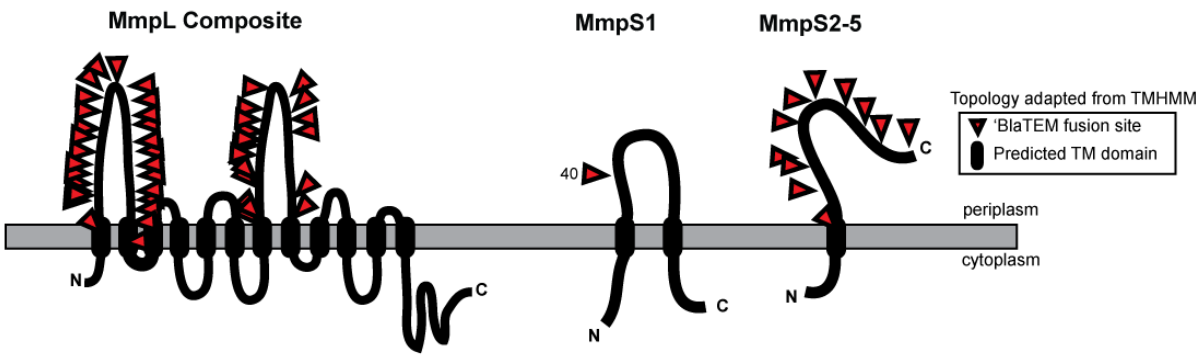

Figure 2.10. EXIT Exported fusions support composite model of MmpL and MmpS proteins. A composite topology was generated by identifying fusions in the two large loops of each of the MmpL proteins. All MmpL proteins identified in EXIT except MmpL10 were identified with exported fusions (red triangles) in the two large loops, consistent with a shared topology for the MmpL family with the two large loops exported to the periplasm. No predicted topology for MmpL10 was consistent with the EXIT identified exported fusion sites, so topology of MmpL10 remains unclear. A composite topology was also generated for MmpS proteins by analysis of EXIT exported fusion sites in the C-terminal domain of MmpS2-5, and the loop of MmpS1.

transmembrane domains of MmpS1 to the periplasm (Figure 2.10). This topology is consistent with a recent study identifying interactions between the C-terminus of MmpS4/5 proteins and the first loop of MmpL4/5 proteins (Wells *et al.*, 2013), both of which are predicted by EXIT to localize to the periplasm.

Emb proteins

Ethambutol (EMB) is a frontline anti-tuberculosis drug known to target mycobacterial cell wall synthesis. EMB resistance mutations are linked to expression and structure of mycobacterial arabinosyl transferases (EmbA,B,C proteins) (Telenti *et al.*, 1997) that function in the biosynthesis of arabinogalactan and lipoarabinomannan, key components in the mycobacterial cell wall (Jackson *et al.*, 2013). Like MmpL3, Emb proteins are essential for *M. tuberculosis* during *in vitro* growth (Goude *et al.*, 2008), important to synthesis of the mycobacterial cell wall, and play important roles in *M. tuberculosis* drug resistance. Despite their significance to mycobacterial cellular physiology and drug resistance, the topology of *M. tuberculosis* Emb proteins has yet to be investigated experimentally. Emb proteins possess approximately 13 predicted transmembrane domains, a large N-terminal loop, and a large C-terminal domain (Krogh *et al.*, 2001). Prediction programs suggest that the topology of EmbB may be different than EmbA and EmbC (Krogh *et al.*, 2001). However, analysis of the 25 unique exported fusion sites in EmbA, B, and C proteins identified in EXIT is consistent with a similar topology for all three proteins, with the N-terminal loop and C-terminal domain localized to the periplasm (Figure 2.11). This topology is in agreement with the published topology of the single Emb orthologue of *Corynebacterium glutamicum*, an acid fast organism related to mycobacteria, that was determined when heterologously expressed in *Escherichia coli* (Seidel *et al.*, 2007).

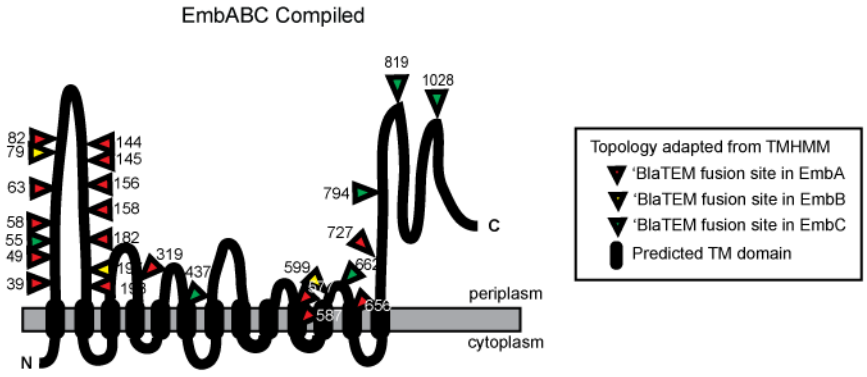

Figure 2.11. EXIT exported fusions clarify topology of EMB proteins important to drug resistance in *M. tuberculosis*. Exported fusion sites in EmbABC proteins in *M. tuberculosis* were analyzed in combination with TMHMM predicted transmembrane domains to generate a composite family topology. Exported fusions in EmbA are depicted in red, EmbB in yellow, and EmbC in green.

ESX (Type VII Secretion System) components

M. tuberculosis encodes five Type VII secretion system islands, named ESX-1 to ESX-5 for the first known exported substrate, ESAT-6. ESX secretion is required for full virulence of *M. tuberculosis*, and deletion of the ESX-1 locus is one of the main attenuating mutations in the BCG vaccine strain (Ligon *et al.*, 2012). ESX gene clusters display a high level of conservation, and encode ESX conserved components (Ecc), ESX specific proteins (Esp), and mycosins (MycP) which form the export machinery. The export mechanism of ESX systems is being actively investigated; however, all ESX systems are believed to form multi-protein transport complexes with components in the cytoplasm and inner membrane (Houben *et al.*, 2014). It is proposed that substrates are targeted to an ESX membrane complex consisting of four conserved membrane proteins (EccBCDE), all of which are required for ESX secretion; however, the individual role of each protein within the complex remains unclear (Houben *et al.*, 2014). EccBCDE and MycP have been localized to the cell envelope in *M. marinum* (Houben *et al.*, 2012). However, the topology for these five membrane proteins has yet to be determined experimentally. EXIT obtained 33 unique exported fusions in EccB, EccD, and MycP family members, representing components of four of the five Type VII secretion systems in *M. tuberculosis* (ESX-1, ESX-2, ESX-3, and ESX-5). Sites of exported fusions localized the large C-terminal domain of EccB and the large N-terminal globular domain of MycP to the periplasm (Figure 2.12). This data supports the proposed topology of EccB and MycP based on *in silico* modeling (Houben *et al.*, 2014). Localization of the globular domain for MycP within the periplasm is significant because MycP has protease activity, and EXIT data demonstrates that the known active site for protease cleavage from crystallography data is localized to the periplasm (Ohol *et al.*, 2010; Wagner *et al.*, 2013). While the function of EccB remains unclear the finding

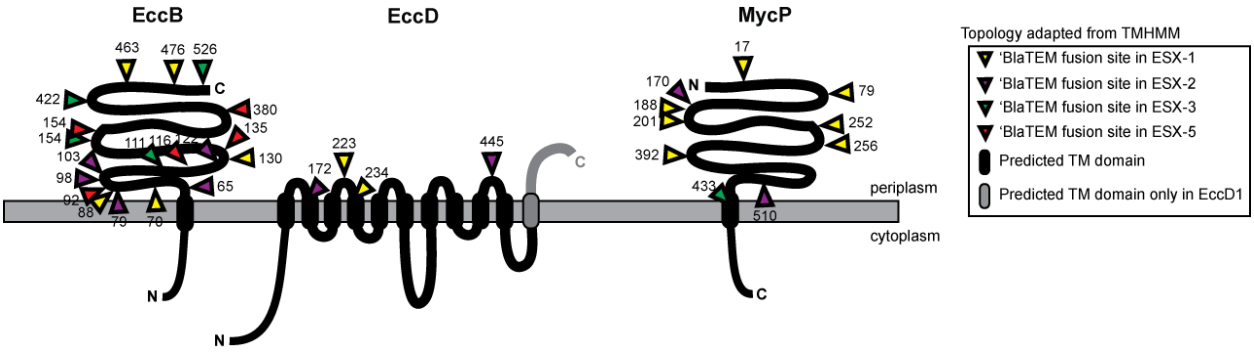

Figure 2.12. EXIT exported fusions clarify topology of ESX Type VII secretion system membrane proteins. Genomic clusters encoding ESX or Type VII secretion systems encode five known membrane proteins thought to compose the ESX secretion machinery. EXIT identified exported fusion sites in three components of these systems, EccB, EccD, and MycP for four of the five ESX secretion systems in *M. tuberculosis*. Transmembrane domain predictions were made by TMHMM and EXIT exported fusion sites were combined to generate a topology diagrams. Yellow triangles represented fusions in ESX-1 membrane proteins, purple in ESX-2, green in ESX-3, and red in ESX-5. The EccD protein in ESX-1 was predicted to have one more transmembrane domain than other EccD proteins, and this extra predicted transmembrane domain was depicted in grey.

that the majority of the protein is localized in the periplasm can help guide future functional studies.

The EccD proteins are predicted to have approximately 10 transmembrane domains (the ESX-1 member of the EccD family, EccD1, is predicted to have 11), with a large N-terminal domain and a large loop after the sixth transmembrane domain (Krogh *et al.*, 2001). Because of their multiple transmembrane domains, EccD proteins are the best candidates for being the membrane channel used for transport (Ligon *et al.*, 2012). However, the function of the EccD family proteins in ESX secretion has not been studied. EXIT did not obtain any exported fusions in the N-terminal domain or the large loop; however, exported fusions were identified in several small loops opposite of these domains (Figure 2.12). This data suggests that the large domains of EccD are cytoplasmic, which will be useful for further study of the function and mechanism of EccD proteins to ESX secretion.

Identification of proteins exported significantly more in vivo than in vitro

Our original hypothesis was that a subset of *M. tuberculosis* proteins are exported significantly more during *in vivo* infection than during *in vitro* growth. One mechanism by which a protein could be exported more *in vivo* than *in vitro* is if its expression levels are induced *in vivo* due to transcriptional or posttranscriptional effects. Several studies have identified genes in *M. tuberculosis* whose expression is upregulated in the host environment (Talaat *et al.*, 2004; Dubnau & Smith, 2003; Rohde *et al.*, 2007b; Timm *et al.*, 2003; Talaat *et al.*, 2007; Rachman *et al.*, 2006b; Schnappinger *et al.*, 2003). Alternatively, the responsible protein export system could be upregulated or only active in the host, resulting in substrates exported more *in vivo* than *in vitro*. There are examples of protein export systems of other bacterial pathogens that are turned on or induced during infection. For example, the *Salmonella* type III secretion system T3SS2 is

only expressed and active following cellular invasion, when exported effector proteins are required for maturation of the *Salmonella* containing vacuole (Malik-Kale *et al.*, 2011). Similarly, the *Yersinia* T3SS is not normally expressed during *in vitro* growth; however, expression can be induced *in vitro* by temperature and calcium concentrations that mimic conditions encountered during infection (Dewoody *et al.*, 2013). In this thesis we will use the terminology "*in vivo* induced" to refer to either category of proteins that are exported more *in vivo* versus *in vitro*. Regardless of mechanism, exported proteins that are induced *in vivo* are attractive candidates for being previously overlooked proteins with critical roles in *M. tuberculosis* virulence.

In an attempt to identify *in vivo* induced exported proteins, the EXIT library bacteria surviving β-lactam treatment in mice were plated in parallel on regular agar media and β-lactam containing agar media (Figure 2.13). The clones recovered on regular media reported on the total population of fusion proteins exported during infection. As described above, these fusions were sequenced leading to the identification of 593 proteins as exported during infection. Clones that grow on media containing β-lactams express fusions that are additionally expressed and exported under *in vitro* conditions. The fusions recovered on β-lactam containing media were sequenced from duplicate experiments, and there was a high degree of correlation between replicates (Figure 2.14A).

Fusion proteins that enabled survival in a β-lactam treated mouse *in vivo* and on β-lactam containing agar *in vitro* are considered examples of proteins that are exported both *in vitro* and *in vivo*. In comparison, any clones that expressed reporter fusions to proteins that are only expressed/exported during infection would survive β-lactam treatment *in vivo* but be unable to grow *in vitro* on media containing β-lactam antibiotics. The majority of *in vivo* exported proteins

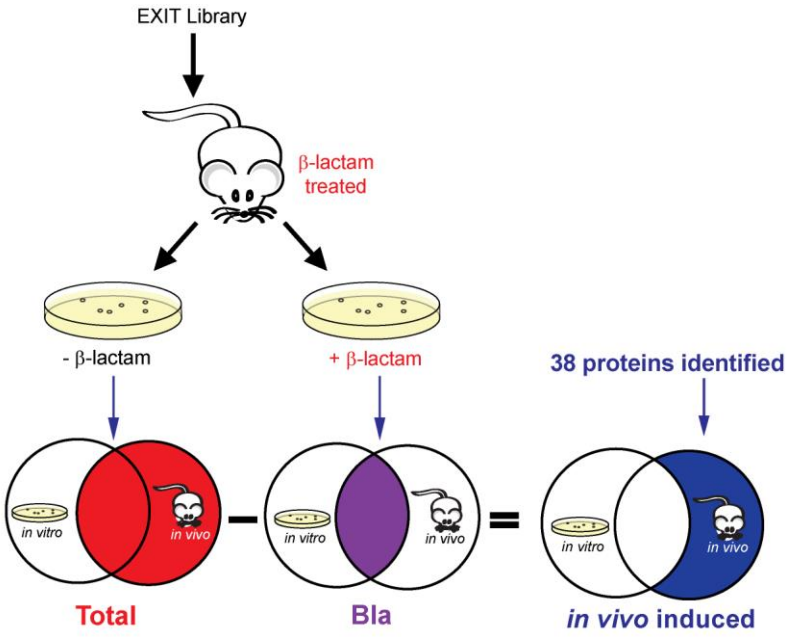

Figure 2.13. Strategy for identification of proteins exported more *in vivo* than *in vitro*. Identification of proteins with export upregulated during infection used a modified EXIT approach. As described in Figure 2.2, mice were infected with the EXIT library, treated with β-lactam antibiotics, and spleens were harvested and homogenized two weeks after infection. Spleen homogenates were plated in parallel on solid agar media to recover all clones (red venn diagram), and solid agar media containing β-lactam antibiotics to recover clones exporting 'BlaTEM fusion proteins during *in vitro* growth (purple venn diagram). The population of clones only identified, or identified in significantly greater abundance on media lacking β-lactams represented proteins whose export was upregulated during infection (blue). Statistical analysis identified 38 proteins.

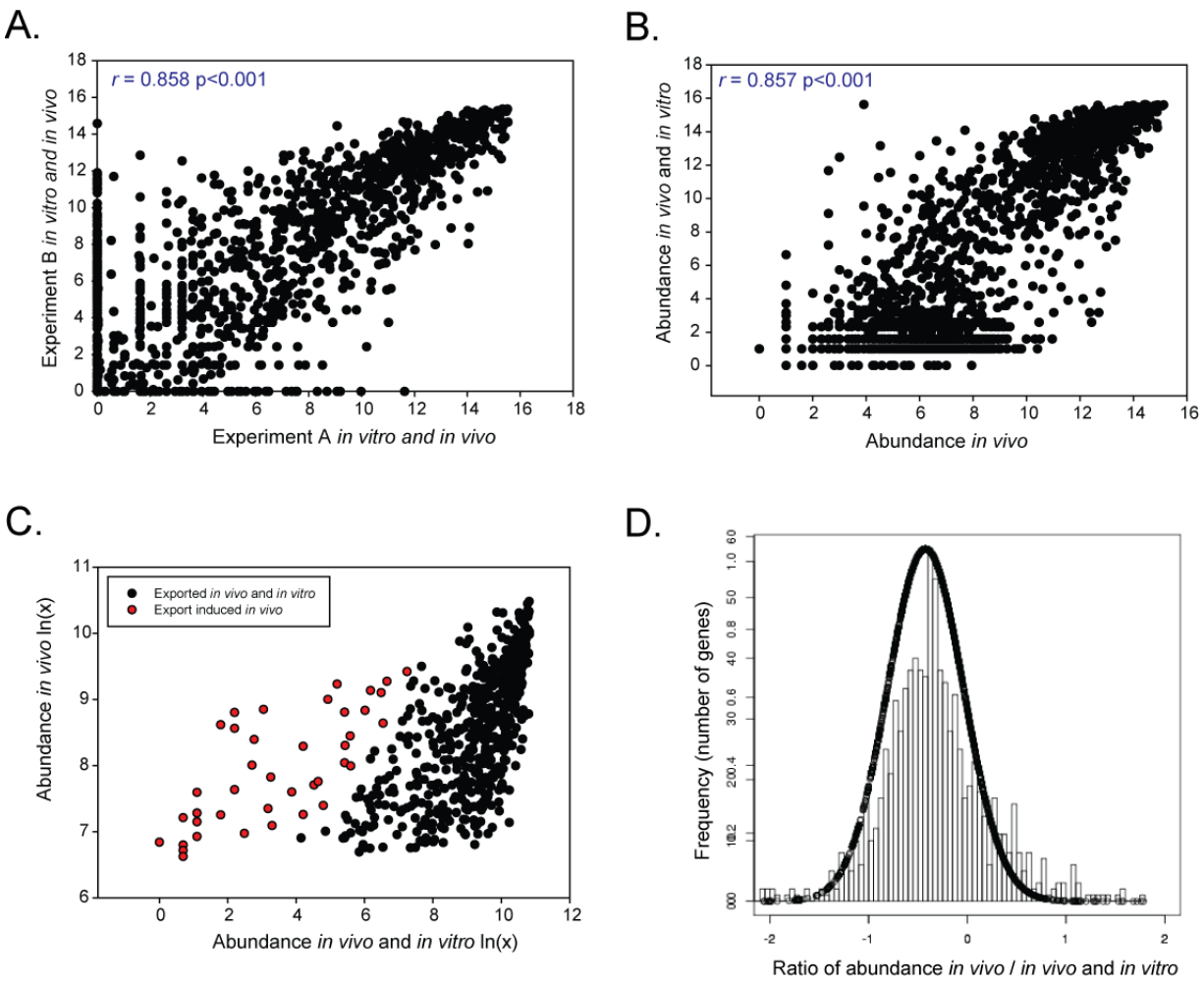

Figure 2.14. Statistical modeling identified 38 proteins exported significantly more *in vivo* than *in vitro*. A. Raw sequenced read count values in the mouse output after β-lactam selection *in vivo* and *in vitro* for each fusion junction site were plotted for replicate experiments A and B. A Pearson Product Moment correlation identified a significant correlation *r* value of 0.858. **B.** Raw sequenced read count values for each fusion junction site were plotted to compare samples with only β-lactam selection *in vivo*, or those that were treated with β-lactams *in vivo* and *in vitro*. A Pearson Product Moment correlation identified a significant correlation *r* value of 0.857. **C.** Raw sequenced read count values for the 593 proteins identified as exported in EXIT were plotted to compare β-lactam treatment *in vivo*, with dual β-lactam treatment *in vivo* and *in vitro*. The majority of proteins identified as exported *in vivo* remained highly abundant after additional β-lactam treatment *in vitro* (black), representing most proteins which are exported similarly *in vitro* and *in vivo*. Genes highlighted in red were identified as statistically less abundant after *in vitro* β-lactam selection, representing proteins exported significantly more *in vivo* than *in vitro*. **D.** Abundance was determined individually for each gene based on sequenced read counts for the most highly abundant fusion site in each experiment. A ratio was determined for abundance after *in vivo* β-lactam treatment compared with abundance after dual β-lactam treatment *in vivo* and *in vitro*. Higher numbers correspond to genes that survived β-lactam treatment in mice but were killed during *in vitro* β-lactam treatment. Statistics were used to identify 38 genes on the right shoulder that were significantly different than the normal population. These genes encoded proteins exported significantly more *in vivo* than *in vitro*.

identified were also exported *in vitro* as shown by a high level of correlation in the abundance of clones recovered on media containing or lacking β-lactams r=0.857 (Figure 2.14B).

As a starting point for comparing the abundance of each fusion after recovery on regular media (*in vivo* exported proteins) versus recovery on media containing β-lactam antibiotics (*in vivo* and *in vitro* exported proteins), we calculated an abundance ratio between the two conditions. The distribution of abundance ratios was uni-modal and normal, with a small right shoulder (Figure 2.14C). The null hypothesis was disproven, and a statistically significant difference was found for a subset of clones recovered on regular media versus those recovered on β-lactams containing media. Using a false discovery rate (FDR) set at 5%, after a correction for multiple comparisons, 38 genes (6% of the total 593 identified by EXIT) were identified as having significantly reduced abundance on β-lactam containing agar *in vitro* compared to regular media, and thus predicted to encode fusions that are exported significantly more *in vivo* than *in vitro* (Figure 2.14D, Table 2.10).

The exported proteins we identified as *in vivo* induced may represent a spectrum of effects, with some exported proteins being induced during infection and others being exclusively exported during infection. Consistent with this range of *in vivo* induced effects, five of the 38 *in vivo* induced exported proteins are predicted to be essential during *in vitro* growth by saturating mutagenesis approaches, meaning they most likely have a function during *in vitro* growth (Griffin *et al.*, 2011; Sassetti *et al.*, 2003). Thus, these proteins likely represent examples of proteins that are exported and functional during *in vitro* growth and induced during infection. Of the 38 *in vivo* induced proteins identified, 14 (37%) are known to be transcriptionally induced during infection (Dubnau *et al.*, 2005; Dubnau *et al.*, 2002; Schnappinger *et al.*, 2003; Rachman *et al.*, 2006b ; Talaat *et al.*, 2007; Rohde *et al.*, 2007a; Rohde *et al.*, 2007b; Sharma *et al.*, 2006;

Srivastava *et al.*, 2007; Rodriguez *et al.*, 2013). The precedent for these proteins being reported as *in vivo* induced provides strong support for the EXIT methodology to identify *in vivo* induced exported proteins. The mechanism of regulation for the remaining 23 proteins remains to be determined.

Among the 38 *in vivo* induced exported proteins are proteins with functions in regulation, host defense, myco-membrane synthesis, and nutrient acquisition. 13 of the *in vivo* induced exported proteins (34%) have demonstrated or predicted roles in virulence (Reddy *et al.*, 2013; Danelishvili *et al.*, 2010; Chuang *et al.*, 2015; Papavinasasundaram *et al.*, 2005; Rifat *et al.*, 2014; Tischler *et al.*, 2013; Marjanovic *et al.*, 2010; McCann *et al.*, 2011; Gioffre *et al.*, 2005; Senaratne *et al.*, 2008; Dutta *et al.*, 2010; Hu *et al.*, 2010; Sassetti & Rubin, 2003; Stewart *et al.*, 2005; Rengarajan *et al.*, 2005). 21 of the 38 proteins (55%) have unknown function. Most of the function unknown proteins are not previously studied and they represent potentially unidentified and underappreciated virulence factors. Additionally, eight (21%) of the 38 proteins do not possess *in silico* predicted export signals, suggesting they may be exported by unique mechanisms. Below we review some of the most notable *in vivo* induced exported proteins identified by the EXIT strategy.

Regulation

Two regulatory proteins were identified as being exported significantly more *in vivo* than *in vitro*. SenX3 is the sensor histidine kinase of the SenX3-RegX3 two component regulatory system implicated in phosphate responsive gene regulation (Rifat *et al.*, 2009; Rifat & Karakousis, 2014; Glover *et al.*, 2007). SenX3 is required for growth in phosphate limiting conditions, and important to virulence (Rifat *et al.*, 2009; Glover *et al.*, 2007; Rifat *et al.*, 2014; Tischler *et al.*, 2013). PknH is a serine/threonine protein kinase unique to pathogenic

mycobacterial strains and known to regulate production of phthiocerol dimycoceroserate (PDIM) (a complex lipid with a role in virulence) and expression of the dormancy regulon (Sharma *et al.*, 2006; Gomez-Velasco *et al.*, 2013; Chao *et al.*, 2010). Deletion of *pknH* results in hypervirulence, as the *pknH* mutant bacteria replicate to a higher bacterial load in mice than WT *M. tuberculosis* (Papavinasasundaram *et al.*, 2005). Both SenX3 and PknH are membrane proteins with predicted transmembrane domains (Krogh *et al.*, 2001).

For these regulators identified by EXIT as *in vivo* induced, their induction is likely a reflection of a transcriptional effect during infection. Transcript levels of *pknH* are upregulated during intracellular growth in macrophages (Sharma *et al.*, 2006; Srivastava *et al.*, 2007). Further, while transcript levels of *senX3* have not been directly tested during infection, *senX3* expression is induced by phosphate depletion (Rifat *et al.*, 2014; Rifat & Karakousis, 2014). Identification of SenX3 as an *in vivo* induced exported protein suggests that *M. tuberculosis* may encounter phosphate limiting conditions during infection, similar to many other bacterial pathogens (Lamarche *et al.*, 2008).

Myco-membrane synthesis

Mycobacteria have a unique outer membrane structure, called the myco-membrane, which is composed of specialized lipids important to cellular physiology and virulence (Figure 1.2). Biogenesis of the myco-membrane requires a series of biosynthetic enzymes and transporters, many of which are unique mycobacterial exported proteins. Two specialized lipid transporters (DrrC and MmpL8) important to myco-membrane synthesis were identified by EXIT as being exported significantly more *in vivo* than *in vitro*. DrrC is an exported membrane protein important for transport of PDIM to the myco-membrane (Camacho *et al.*, 2001; Camacho *et al.*, 1999; Forrellad *et al.*, 2013b). MmpL8 is a multi-membrane spanning integral membrane

protein that plays a role in biosynthesis and transport of sulfolipid-1 (SL-1), an abundant lipid in the myco-membrane (Domenech *et al.*, 2004; Converse *et al.*, 2003). Previous studies are consistent with PDIM levels and *mmpL8* expression increasing during infection (Jain *et al.*, 2007; Rodriguez *et al.*, 2013). Our identification of DrrC and MmpL8 being exported at increased levels during infection may indicate a need for increased levels of specialized lipids within the myco-membrane during infection.

Host Defense

Two of the proteins identified as *in vivo* induced exported proteins play roles in defense from and manipulation of the host: Mmco and Rv3654c. Mmco is a periplasmic multi-copper oxidase, and one of several proteins that provide protection from copper toxicity, a newly described macrophage defense strategy (Rowland & Niederweis, 2013; Shi & Darwin, 2015). The expression of *mmco* is controlled by a copper responsive regulator, RicR (Festa *et al.*, 2011), and Mmco protein levels increase after incubation with copper (Shi *et al.*, 2014). Our identification of Mmco as exported more *in vivo* than *in vitro* confirms previous studies suggesting that *M. tuberculosis* encounters high copper concentrations during murine infection (Talaat *et al.*, 2004; Ward *et al.*, 2010; Neyrolles *et al.*, 2015).

A particularly interesting protein on the list of *in vivo* induced exported proteins is Rv3654c, which has a function suppressing macrophage apoptosis during infection (Danelishvili *et al.*, 2010). *In silico* predictions do not predict Rv3654c to be an exported protein, and thus it is also on our list of 32 proteins identified by EXIT with no *in silico* export signal. However, Rv3654c was previously identified in macrophage lysates cleared of intracellular *M. tuberculosis*, suggesting it is exported out of the bacterial cell and into the macrophage during infection (Danelishvili *et al.*, 2010). Our EXIT results provide important confirmation of

Rv3654c being an exported protein. Further, the identification of Rv3654c on our list of exported proteins induced *in vivo* is consistent with previously described transcriptional induction of the *rv3654c* gene during macrophage infection (Danelishvili *et al.*, 2010). Interestingly, *rv3654c* is located within an operon encoding proteins with homology to tight adherence (Tad) secretion system components (Tomich *et al.*, 2007; Danelishvili *et al.*, 2010), offering a potential export mechanism for this unconventional exported protein. Future studies are required to determine the contribution of Tad-like proteins to export of Rv3654c.

Nutrient acquisition

Seven (18%) of the 38 proteins EXIT identified as exported significantly more *in vivo* than *in vitro* are predicted to play a role in nutrient acquisition, reflecting altered nutritional requirements or availability during infection, compared to *in vitro* growth conditions. Iron, sulfur, and nitrogen are all fundamental inorganic nutrients required for life and examples of elements that bacteria must acquire to successfully grow in an intracellular environment. We identified multiple examples of proteins important for acquisition of these elements as being export induced during infection. Iron acquisition is a well-studied virulence mechanism for bacterial pathogens (Becker & Skaar, 2014). Biosynthesis of siderophores, iron chelating molecules designed to compete for and deliver host iron to bacteria, is induced during infection (Timm *et al.*, 2003; Schnappinger *et al.*, 2003). We identified MtbE, one essential component of the siderophore biosynthetic operon, as being induced during infection, which is consistent with transcriptional induction stimulated by low iron conditions *in vivo* (Reddy *et al.*, 2013). NarK2 and NarK3 are two of four predicted nitrate/nitrite transporters in *M. tuberculosis*, with NarK2 being shown to be a H^+:nitrate antiporter (Giffin *et al.*, 2012). Expression of *narK2* is induced in hypoxic conditions and during infection, and expression and nitrate transport occurs concurrently

with increased nitrate reductase activity, yielding large quantities of nitrite (Sohaskey & Wayne, 2003; Shi *et al.*, 2005). Although the assumption may have been that induction of *narK2* is required for acquisition of nitrogen for nutrition during infection, nitrate imported by *narK2* is thought to be used primarily as a terminal electron acceptor for respiration rather than assimilation (Sohaskey & Wayne, 2003; Gouzy *et al.*, 2014). Thus, the source of nitrogen as a nutrient, the pathways responsible for nitrogen acquisition and assimilation during infection, and whether nitrogen acquisition pathways are required for virulence remain to be determined (Gouzy *et al.*, 2014). Rv1739c is one of three predicted SulP family sulfate transporters in *M. tuberculosis*, and it is shown to transport sulfate when expressed in *E. coli* (Zolotarev *et al.*, 2008). However, the role of SulP family sulfate transporters to sulfate transport *in vitro* or during infection is currently unknown. Our identification of Rv1739c as exported *in vivo* suggests that this understudied group of proteins could be important during infection (Niederweis, 2008; Zolotarev *et al.*, 2008).

Mce lipid transporters

M. tuberculosis encodes four *mce* operons predicted to encode multi-protein complexes that import lipids (Casali & Riley, 2007). Mce4 is shown to import cholesterol (Pandey & Sassetti, 2008; Mohn *et al.*, 2008), and recent evidence suggests that Mce1 plays a role in recycling mycolic acids, which are central components of the myco-membrane (Forrellad *et al.*, 2014; Cantrell *et al.*, 2013). In contrast, the substrates for Mce2 and Mce3 transporter systems have yet to be characterized, although Mce2 and Mce3 are known to contribute to *M. tuberculosis* virulence (Marjanovic *et al.*, 2010; McCann *et al.*, 2011; Gioffre *et al.*, 2005; Senaratne *et al.*, 2008; Dutta *et al.*, 2010). The operons encoding Mce2 and Mce3 appear to be repressed during *in vitro* growth (Santangelo *et al.*, 2002; Santangelo Mde *et al.*, 2009; Forrellad

et al., 2013a; Gioffre *et al.*, 2005) but expressed during infection (Santangelo *et al.*, 2002; Schnappinger *et al.*, 2003). Consistent with these findings, several exported components of Mce2 and Mce3 transporters were found to be significantly induced during infection (Mce2E, YrbE3B, and Mce3D). Direct investigation of the suite of exported components of the Mce2 and Mce3 systems revealed a consistent trend in being exported more *in vivo* than *in vitro*, although some of these proteins did not meet our stringent statistical cutoffs of significance (Figure 2.15).

Eighteen percent of proteins identified by EXIT as exported more *in vivo* than *in vitro* belonged to the category of proteins required for nutrient acquisition, underscoring the differences in nutrient availability and acquisition between *in vitro* and *in vivo* growth. Many of these nutrient acquisition systems are understudied, and the nutrients they are responsible for acquiring have yet to be determined, highlighting the limitations of our knowledge about growth *in vivo*.

In vivo induced exported proteins of unknown function

One goal of EXIT was to provide information on function unknown proteins. Determining a given protein's localization and regulation are two common strategies for learning about a protein of unknown function. Towards this goal, EXIT experimentally identified 21 proteins of unknown function as being exported and induced *in vivo*. These proteins represent the most exciting category of proteins identified by EXIT. Two of these 21 proteins belong to the MmpL/S family of mycobacterial proteins, MmpS2 and MmpL13a. Like other MmpL/S family proteins these membrane proteins are likely to be transporters; however, they remain to be characterized (Domenech *et al.*, 2005).

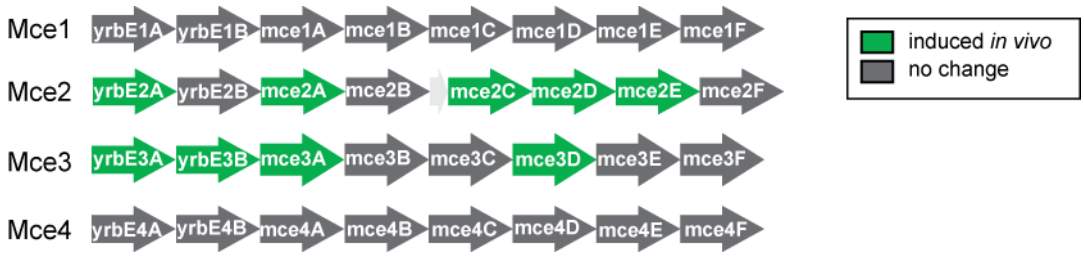

Figure 2.15. Proteins of the Mce2 and Mce3 lipid transporters of *M. tuberculosis* are reliably exported more *in vivo* than *in vitro*. Genes in *mce* operons encoding multi-protein Mce lipid transporters were analyzed for evidence of being exported more *in vivo* than *in vitro*. While only three genes were statistically significant (*mce2E*, *yrbE3B*, and *mce3D*) multiple other genes in *mce2* and *mce3* operons were depleted more than two fold change after *in vitro* β-lactam treatment (green arrows). In contrast, genes in *mce1* or *mce4* operons showed no change in abundance after *in vitro* β-lactam treatment (grey arrows), demonstrating export *in vitro* as well as *in vivo*.

Three of the *in vivo* induced exported proteins of unknown function have no *in silico* predicted export signal: PPE54, PPE60, and Rv3707c. PE/PPE proteins are surface localized, and expression of *ppe* genes varies widely in different *in vitro* conditions (Voskuil *et al.*, 2004; Banu *et al.*, 2002; Brennan *et al.*, 2001). Neither PPE54 or PPE60 have a YxxxD/E motif, described as being important for Type VII secretion of PE/PPE proteins (Daleke *et al.*, 2012), and neither protein is located in an operon with a Type VII secretion system. Consequently, the mechanism of export of these proteins is currently unknown. Expression of *ppe54* is induced during human infection (Rachman *et al.*, 2006b), and PPE54 appears to play a role in preventing phagosomal maturation, an important process in *M. tuberculosis* virulence (Brodin *et al.*, 2010). Rv3707c is important for preventing phagosomal maturation in the attenuated vaccine strain *Mycobacterium bovis* BCG, and for growth of BCG in macrophages (Stewart *et al.*, 2005). The mechanism of export for Rv3707c is also unknown because it possesses no *in silico* predicted export signal. Importantly, in this thesis we confirmed that Rv3707c was exported to the cell wall in *M. tuberculosis* (Figure 2.8B).

Potentially the most interesting category of unknown function proteins we identified as *in vivo* induced exported proteins are those that are predicted to be fully secreted. Of all the categories of exported proteins, the fully secreted proteins are most likely to localize outside of the bacterial cell and interact with the host. Two proteins, Rv0559c and Rv0817, have predicted Sec signal peptides and, because they lack any additional predicted transmembrane domains, they may be fully secreted. Both of these proteins have been identified as core mycobacterial proteins, defined as proteins with homologues throughout mycobacteria but no homologues outside of actinomycetes (Marmiesse *et al.*, 2004). The potential role of Rv0559c and Rv0817 in virulence has yet to be tested, but these proteins are particularly interesting virulence candidates.

Future studies will help to understand how and why *M. tuberculosis* regulates export of these proteins, and their potential role in virulence and immunogenicity. Identifying these proteins as exported and further characterization of their role during disease could lead to a better understanding of *M. tuberculosis* infection and host-pathogen interactions.

Discussion

EXIT is the first genome-wide method for discovering bacterial proteins exported during infection. As such, it represents a powerful tool for identifying new proteins with roles in host-pathogen interactions. Previous methodologies used to identify exported proteins of bacteria were limited to identifying proteins exported during growth in lab media, *in vitro*. With *in vitro* based systems, potentially interesting and important proteins exported only *in vivo* in an animal model of disease would have been overlooked.

Besides EXIT there are recent studies of bacterial protein export *in vivo*; however, so far, these efforts are limited to direct testing of individual proteins. In several cases the studies utilize individual protein fusions to the ‘BlaTEM reporter, however, the fusions were detected using a cell-permeable fluorescent β-lactamase substrate (CCF2-AM) to monitor transport out of the bacterium and into host cells during infection (Pechous *et al.*, 2013; Danelishvili *et al.*, 2014; Ge *et al.*, 2009; Broms *et al.*, 2012). While this approach can be used to identify certain types of proteins exported in the context of host cells, it does not take advantage of ‘BlaTEM as a selectable marker and screening of individual clones is required. Thus, scaling up to a level of genome-wide discovery, as done in EXIT, would be difficult with this method. In a different approach, click-chemistry was recently used to study export of Type III secretion system (T3SS) proteins exported during infection (Mahdavi *et al.*, 2014). This approach uses an engineered t-

RNA to incorporate an unnatural amino acid (azidonorleucine, ANL) into proteins produced by *Yersinia entercolitica* (Tanrikulu *et al.*, 2009; Mahdavi *et al.*, 2014). ANL can be targeted by click-chemistry to selectively enrich for ANL-containing proteins, addressing one of the difficulties of using mass spectrometry to identify exported bacterial proteins amongst a mixture of predominately host proteins. The authors used this strategy to confirm the export and determine the timing of export for T3SS substrates injected by extracellular *Y. entercolitica* into cultured macrophages (Mahdavi *et al.*, 2014). However, the authors were unable to conclusively identify any new exported proteins; the other proteins identified in this study are likely to be cytoplasmic contaminants (Mahdavi *et al.*, 2014). Further refinement of the ANL methodology will be required before this method is broadly applicable to identifying *in vivo* exported proteins on a large-scale.

The EXIT Strategy

The ‘BlaTEM reporter and EXIT methodology are theoretically compatible with any bacterium that is either naturally, or can be made genetically, β-lactam sensitive. This study focused on application of EXIT to the study of *M. tuberculosis* and identification of *in vivo* exported proteins during acute murine infection. However, EXIT could be used in the future to study the *in vivo* exported proteome for a variety of pathogens, in diverse models of infection, and could lead to a better understanding of host-pathogen interactions for multiple bacterial diseases.

As a new method, there were several obstacles to overcome in the development of EXIT. We were initially concerned about the potential of having a high background of non-exported fusions due to trans-complementation and/or lysis of bacteria expressing non-exported fusions, which could release active β-lactamase and potentially promote the survival of surrounding

bacteria. However, proof of principle experiments in which *in vivo* β-lactam selection was tested with mixed populations of β-lactam sensitive and β-lactam resistant clones showed significant enrichment of β-lactam resistant clones, indicating that any background that possibly exists was insufficient to mask the power of the β-lactam selection (Figure 2.4, Table 2.4). We also had to optimize β-lactam dosage and route of infection to be compatible with screening the large EXIT library. Fortunately several different routes of infection are routinely used to model tuberculosis in mice (Orme & Gonzalez-Juarrero, 2007). A low dose aerosol model is most similar to natural infection, where humans are thought to be infected by inhalation of as few as one bacterium (Russell *et al.*, 2010). However, a low aerosol dosage (100 bacteria) is not compatible with screening a large library and we were unsuccessful in our effort to establish conditions to achieve a sufficiently high pulmonary dose (~50,000 cfu/lungs) for surveying the EXIT library in a comprehensive manner. These difficulties in achieving a high lung burden by the aerosol route of infection explain why genome-wide saturating mutagenesis screens for *M. tuberculosis* (e.g. TraSH, DeADMAn) use mice that are infected by intravenous injection and analyzed the library recovered from the spleens (~ 1 x10^6 cfu/spleen) (Sassetti & Rubin, 2003; Zhang *et al.*, 2013; Lamichhane *et al.*, 2005). Despite the focus on spleens instead of the lungs, these genome-wide screens have effectively identified a large number of virulence factors later confirmed to be important to growth and survival in both spleens and lungs (Sassetti & Rubin, 2003; Forrellad *et al.*, 2013b; Zhang *et al.*, 2013; Lamichhane *et al.*, 2005). Thus, for EXIT we similarly utilized mice that were infected by intravenous injection and comprehensively surveyed the library in spleens. We also gathered and analyzed data from lungs to identify trends; however, the reported results of EXIT focus primarily on the robust dataset from the spleens.

Proteins identified as exported *in vivo*

EXIT identified 593 proteins as exported *in vivo*. Not only is this an impressive accomplishment because it is the first effort to identify proteins on a genome-wide level as exported *in vivo*, but it also represents a large number of exported proteins being identified at one time by a genetic reporter in a bacterial pathogen. The effectiveness of EXIT was a result of the following features of the system: 1) the highly comprehensive library (99% of the genome represented with at least one in frame fusion in the library), 2) the use of the 'BlaTEM reporter as a selectable marker, and 3) the use of next-generation sequencing to identify enriched exported fusions. EXIT identified proteins known to localize within each exported compartment: cytoplasmic membrane (e.g. MmpL), periplasm (Mmco), cell wall (penicillin binding proteins, Ag85A/FbpA), myco-membrane (outer membrane protein A, OmpA), fully secreted proteins (Mpt32, Mpt63, Mpt64), and even proteins shown to be secreted out of the phagosome during intracellular *M. tuberculosis* growth (19kD/LpqH). 100 of the proteins identified by EXIT had not previously been experimentally demonstrated as exported by genetic reporter or mass spectrometry based methodologies (Gu *et al.*, 2003; Rosenkrands *et al.*, 2000a; Xiong *et al.*, 2005; Mawuenyega *et al.*, 2005; Wolfe *et al.*, 2010; Malen *et al.*, 2007; de Souza *et al.*, 2011; Bell *et al.*, 2012; Gunawardena *et al.*, 2013; Braunstein *et al.*, 2000; Gomez *et al.*, 2000; McDonough *et al.*, 2008; McCann *et al.*, 2011; Feltcher *et al.*, 2015).

EXIT identified 32 proteins with no *in silico* predicted export signals and six unannotated regions of the genome that appear to encode exported proteins. Three of the proteins with no predicted export signals were selected for further analysis and all three were successfully validated as exported and localized to the cell wall of *M. tuberculosis* (Figure 2.8B). Given this successful validation, it seems likely that there are additional examples of unconventional

exported proteins on this list of 32 exported proteins lacking *in silico* predicted export signals There is clear precedence for novel export systems in bacterial pathogens, and commonly these specialized export systems play important roles in virulence. However, it is often difficult to identify substrates of these export systems (Ligon *et al.*, 2012). Therefore, we are excited about the potential for proteins identified by EXIT but lacking *in silico* predicted export signals to represent new exported substrates of currently unknown export pathways.

There was an interesting group of eight proteins not identified by EXIT as exported in the spleen but predicted to be exported in the lungs, including four PE_PGRS proteins (Table 2.9). PE_PGRS proteins are a subfamily of PE/PPE proteins that additionally contain polymorphic guanosine-cytosine-rich sequences (PGRS) (Goldberg *et al.*, 2014). Among the 4 PE_PGRS family we identified in the lung, but not the spleen, is PE_PGRS33. PE_PGRS33 is the best studied PE_PGRS protein, and it contributes to growth of the *M. bovis* BCG vaccine strain in macrophages (Brennan *et al.*, 2001), Additionally, PE_PGRS33 may modulate the host cytokine response during *M. tuberculosis* infection (Bottai *et al.*, 2014; Goldberg *et al.*, 2014). The identification of these 4 PE_PGRS proteins in the lung data from EXIT but not in the more comprehensive spleen data is intriguing as it could reflect lung specific mechanisms of expression or export for these proteins.

We had prior evidence that the ‘BlaTEM reporter is compatible with fusions containing large N-terminal segments of the native protein (McCann *et al.*, 2007), and EXIT successfully identified exported fusions throughout many large exported proteins (Figure 2.9B, APPENDIX II). For every protein identified, the site of enriched fusions to the reporter is provided in APPENDIX II. There are many prediction programs available to predict transmembrane domains and membrane protein topology; however, the predictions of these programs often disagree

(Krogh *et al.*, 2001; Sonnhammer *et al.*, 1998; Hofmann & Stoffel, 1993; Claros & von Heijne, 1994; Jones *et al.*, 1994). For example, different prediction programs led to multiple published topology predictions for MmpL3 (Li *et al.*, 2014; Rayasam, 2014; Remuinan *et al.*, 2013; Poce *et al.*, 2013; Tahlan *et al.*, 2012; La Rosa *et al.*, 2012; Sandhu & Akhter, 2015; Varela *et al.*, 2012). Fusion site information from EXIT provided strong evidence in favor of one of the predicted topologies, generated by TopPred (Figure 2.9B) (Claros & von Heijne, 1994). The EXIT database of exported fusion sites for 337 membrane proteins will help in selecting the correct topology prediction and the dataset could also be useful for building an improved prediction algorithm to define the location of extra-cytoplasmic domains.

EXIT identified 57% of the *in silico* predicted exported proteome (1040 proteins) as exported *in vivo* (Krogh *et al.*, 2001; Petersen *et al.*, 2011; Sonnhammer *et al.*, 1998; Sutcliffe & Harrington, 2004; McDonough *et al.*, 2008). *In silico* algorithms remain the easiest and most used method of identifying a protein's localization; however, some of the predictions may be wrong. It is interesting to note that 39% of the *M. tuberculosis* proteins predicted as exported *in silico* but NOT identified by EXIT have never been identified as exported by any method, including mass spectrometry based approaches (Gu *et al.*, 2003; Rosenkrands *et al.*, 2000a; Xiong *et al.*, 2005; Mawuenyega *et al.*, 2005; Wolfe *et al.*, 2010; Malen *et al.*, 2007; de Souza *et al.*, 2011; Bell *et al.*, 2012; Gunawardena *et al.*, 2013; Braunstein *et al.*, 2000; Gomez *et al.*, 2000; McDonough *et al.*, 2008; McCann *et al.*, 2011; Feltcher *et al.*, 2015). This lack of identification by any experimental method could reflect these proteins not being truly exported, being present in very low abundance, or not expressed in the conditions tested. EXIT was limited to testing proteins expressed during the first two weeks (acute phase) of murine infection.

Testing for export during later stages of murine infection or in additional animal models could potentially identify additional exported proteins of *M. tuberculosis.*

Lack of identification by EXIT is not sufficient evidence to conclude a protein is not exported. Genetic reporters for export can be limited by incompatibility with specialized export systems, the level of expression required for a positive export signal (β-lactam resistance), and instability of certain fusion proteins. For example, despite the representation of ESAT-6 like proteins by in frame fusions in the input library, none of these classic ESX secreted proteins were identified as exported by EXIT. This was initially surprising, as a recent paper demonstrated that the ‘BlaTEM reporter could report on ESX secreted proteins; however, this paper has since been fully retracted (2013; Rosenberger *et al.*, 2012). It is currently unknown whether the lack of identifying ESAT-6 like proteins with the ‘BlaTEM reporter is due to incompatibility of this reporter with ESX secretion, instability of reporter fusions to this category of proteins, or interrupted binding of ESAT-6 like protein fusions to their co-secreted proteins. PE_PGRS proteins contain a predicted YxxxD/E motif, and are thought to be exported by ESX secretion (Daleke *et al.*, 2012). Four PE_PGRS proteins were identified by EXIT in the lungs (Table 2.9), suggesting that ‘BlaTEM reporter may not be inherently incompatible with ESX secretion. However, it is notable that three of these four PE_PGRS proteins also had *in silico* predicted Sec signal peptides (Table 2.9) (Petersen *et al.*, 2011), thus it will be important to determine which targeting domain (Sec signal peptide or YxxxD/E motif) is responsible for export of these PE_PGRS proteins.

Proteins identified as exported significantly more *in vivo* than *in vitro*

The original impetus for identifying proteins exported during infection was the hypothesis that some proteins would only be exported, and thus only identified as exported, *in*

vivo. EXIT provided evidence for 38 proteins being *in vivo* induced exported proteins (i.e. proteins that are exported to higher levels during infection), most of which have unknown function (Table 2.10). Determining whether these proteins are uniquely expressed and/or exported during infection will require further study. It is notable that 13 of these proteins (34%) have not been identified previously as exported *in vitro*, including by highly sensitive MS methods, and may represent proteins uniquely exported during infection (Gu *et al.*, 2003; Rosenkrands *et al.*, 2000a; Xiong *et al.*, 2005; Mawuenyega *et al.*, 2005; Wolfe *et al.*, 2010; Malen *et al.*, 2007; de Souza *et al.*, 2011; Bell *et al.*, 2012; Gunawardena *et al.*, 2013; Braunstein *et al.*, 2000; Gomez *et al.*, 2000; McDonough *et al.*, 2008; McCann *et al.*, 2011; Feltcher *et al.*, 2015).

The mechanism for *in vivo* induced export of each protein has yet to be determined, and there are several different possibilities. First, a subset of proteins with induced export during infection will be transcriptionally induced during infection. Secondly, some of these proteins may be post-transcriptionally regulated: translated or stabilized during infection. Finally, we hypothesize that the export of some proteins may be directly controlled, such that the protein is maintained within the bacterial cytoplasm until a signal is received to initiate export. This could occur through regulated expression and/or building of the export apparatus, or regulated triggering of the export process. Future studies will be necessary to define the mechanism of *in vivo* regulation for each of these proteins.

The identification of exported proteins lacking *in silico* predicted export signals and identification of unannotated exported proteins demonstrates the power of an unbiased genome-wide strategy as well as an unbiased sequencing platform. The ability to assign an exported location to unpredicted exported proteins exemplifies how EXIT can provide functional genomic

information to open reading frames (ORFs) of unknown function. Identification of exported proteins is important because exported proteins often are virulence determinants for pathogens (Ligon *et al.*, 2012). Additionally, exported proteins, and protein export in general are potentially good targets for anti-virulence treatments and antibiotics as antibiotics would not have to fully penetrate the bacterial cell to reach their target (Feltcher *et al.*, 2010). Finally, because bacterial exported proteins are more exposed to the host environment they are more readily available for recognition by the host immune system than cytoplasmic proteins. Nearly all of the immunogenic antigens studied for *M. tuberculosis* are exported proteins that were identified by *in silico* prediction algorithms as well as analysis of fully secreted and cell wall fractions of *in vitro* grown *M. tuberculosis* (Boesen *et al.*, 1995; Kurtz & Braunstein, 2005; Singh *et al.*, 2001; Samanich *et al.*, 1998; Samanich *et al.*, 2000; Laal *et al.*, 1997; Lyashchenko *et al.*, 1998). Antigenic proteins in *M. tuberculosis* are being used to develop new targets for diagnostic development as well as new vaccine candidates. Thus, identification of new *in vivo* exported proteins has the added potential value for revealing novel antigens to be considered for diagnostic and vaccine development (Weiner & Kaufmann, 2014; Ivanyi, 2014).

Methods:

Bacterial growth. In this study, we used the bacterial strains listed in Table 2.2, and plasmids as listed in Table 2.1. *M. tuberculosis* strains were grown in Middlebrook 7H9 broth (Difco) supplemented with 1x albumin dextrose saline (ADS), 0.5% glycerol, and 0.025% Tween-80 (7AGT) (Braunstein *et al.*, 2002). As needed, growth media was supplemented with 20 µg/mL kanamycin (Acros), 50 µg/mL hygromycin (Roche), or 50 µg/mL carbenicillin (Sigma). *E. coli*

strains were grown on Luria-Bertani medium (Fisher) supplemented as necessary with 40 μg/ml kanamycin, 150 μg/ml hygromycin, and 100 μg/ml carbenicillin.

Construction of the 'BlaTEM reporter plasmid pDW31. Detailed plasmid information can be found on Table 2.1. pMB219 was digested with Van91I, overhangs were repaired with T4 DNA polymerase to generate blunt ends and the resulting fragment was digested with Nru1 to remove the kanamycin cassette. A hygromycin cassette derived from pYUB412 as a SmaI/EcoRV fragment was then ligated into the Van91I/NruI digested pMB219 to yield pDW1. In order to reduce the size of the plasmid backbone, pDW1 was digested with HpaI and Van91I, blunt ended with T4 DNA polymerase, and religated to obtain pDW10. '*blaTEM*, liberated from pJES110 as a BamHI and ClaI fragment, was next cloned into BamHI and ClaI digested pDW10 to yield pDW14. Several additional steps lead to the introduction of a *'sacB* reporter downstream of the *'blaTEM* reporter to generate the final library vector pDW31. The *'sacB* reporter was intended to be used as a counter-selectable marker; however, the counter-selection proved to be problematic and was not utilized in the final EXIT strategy.

EXIT Library Construction. M. tuberculosis genomic DNA (gDNA) was prepared as previously described (Pavelka & Jacobs, 1999) from the *M. tuberculosis* Δ*blaC* mutant, PM638 (Flores *et al.*, 2005). Genomic DNA fragments were generated by partial digestion with AciI and HpaII over a 2 hour digestion time optimized for generation of fragments between 500bp and 5kb. pDW31 was digested with ClaI, the restriction enzyme site immediately upstream of the 'BlaTEM reporter. The AciI and HpaII DNA fragments were then ligated into the ClaI site with T4 DNA ligase (Invitrogen). Ligated plasmids were transformed into MegaX DH10 Electrocompetent cells (Invitrogen). 5.64×10^6 *E. coli* transformants were pooled and plasmids were isolated using QiaFilter Plasmid Giga Kit (Qiagen). Plasmids isolated from *E. coli* were

transformed into PM638, *M. tuberculosis* H37Rv Δ*blaC*, as previously described (Braunstein *et al.*, 2002). $5.4x10^6$ *M. tuberculosis* transformants from 50 transformations were pooled to generate the final *M. tuberculosis* EXIT library.

Mouse infection with the EXIT library. For identification of *in vivo* exported proteins, 8-10 week old female BALB/c mice were infected with $2x10^6$-$4x10^6$ cfu of the *M. tuberculosis* EXIT library by tail vein injection, as previously described (Braunstein *et al.*, 2003). For each of two separate experiments, 30 mice were infected. One day after infection, lungs, livers and spleens from six mice were harvested to determine initial dose and organ burden. 24 mice per experiment were maintained and treated with β-lactam antibiotics. One day after infection mice began treatment by oral gavage twice daily with 40 mg amoxicillin (MP Biomedicals 190145 or Sigma A8523) and 8mg probenecid (Sigma P8761) administered in 0.25M NaOH in PBS. 14 days post infection mice were CO_2 euthanized and spleens and lungs were harvested to collect surviving bacteria.

Recovery of bacteria that survived in vivo β-lactam treatment. Organ homogenates were plated undiluted onto solid agar media containing 7AGT and cyclohexamide (as described above in *Bacterial growth*). For determining fusions exported both *in vivo* and *in vitro* organ homogenates were plated in parallel onto solid agar media containing carbenecillin (a β-lactam antibiotic). Plates were incubated at 37°C for three weeks, after which colonies were pooled for plasmid DNA isolation.

Isolation of plasmid DNA from recovered M. tuberculosis. Library plasmids were isolated from *M. tuberculosis* using a modification of the QIAfilter Plasmid Maxi Kit (Qiagen). *M. tuberculosis* colonies were pooled from agar plates as described above, and incubated with 1% glycine at 37°C for 24 hours prior to harvest, pelleted at room temperature and frozen at -20°C

overnight. Pellets were resuspended in P1 buffer with RNase and 10 mg/mL lysozyme, and incubated at 37°C for one hour. Buffer P2 was added and mixed by inverting, and incubated 5 minutes. Buffer P3 was added and mixed by inverting. Lysate was poured into Qiafilter cartridge, and incubated 10 minutes. The lysate was filtered through the cartridge, and then filtered twice through 0.22µm Steriflip units (Millipore) to ensure sterility before removal from the BSL-3 facility. Filtered lysate was next added to Qiagen-tips that were pre-equilibrated with Buffer QBT. The column was washed with a 1:1 mixture of chloroform and methanol, washed twice with Buffer QC, and eluted with Buffer QF. DNA was precipitated with isopropanol and centrifugation. DNA pellets were washed with 70% ethanol, dried, and resuspended in TE buffer.

Sample preparation for Illumina sequencing. Detailed primer information can be found on Table 2.3. Plasmids were linearized by DraI digestion and then sheared by nebulization at 45 psi for 3 minutes in a glycerol TE buffer (53% Glycerol, 37mM Tris-HCl, 5.5mM EDTA) as described in NimbleGen Arrays User's Guide: Sequence Capture Array Delivery v3.1, generating fragments ranging from 200-800bp in length. The nebulized DNA was then purified with the QiaQuick PCR Purification kit (Qiagen). Purified DNA was prepared for sequencing as previously described (Long *et al.*, 2015). Fragments were end repaired using Blunting kit (New England Biolabs E1201L), purified using PCR Purification kit and resuspended in DNase free dH_20. Fragments were tailed with dATP (Roche) by Taq polymerase (Invitrogen) for 45 minutes at 72°C. A double stranded adaptor was constructed from Adaptor 1.1 and Adaptor 2.1 primers mixed with $MgCl_2$ and incubated in a thermocycler at 1% ramp from 95°C to 20°C. Double stranded adaptors were ligated to the sheared plasmid fragments with T4 DNA ligase (Invitrogen) and incubated at 16°C overnight. Using the adaptor ligated DNA as template,

amplification of ‘*blaTEM* fusion junctions was performed with adaptor and ‘*blaTEM* targeting primers (‘Bla IL, ‘Bla IL 3b, ‘Bla IL 4b, ‘Bla IL 5b and Adaptor 1 (containing multiplex barcode 1), Adaptor 2 (containing multiplex barcode 2), and Adaptor 3 (containing multiplex barcode 3)) with Phusion High Fidelity Polymerase (New England Biolabs). 200-400 bp fragments were isolated from the PCR amplification and purified with QiaQuick Gel Purification kit (Qiagen). Samples were sequenced using a HiSeq (Illumina) generating paired end multiplexed sequencing reads.

Computational analysis of sequencing data to identify ‘blaTEM reporter fusion sites. Paired end reads included adapter regions containing two barcodes: a multiplexing barcode, and a random barcode (NNANNANNN), as previously described with minor modifications (Long *et al.*, 2015). The reads were first de-multiplexed. The adapter regions were identified using custom R code which used the Bioconductor biostrings package (H. Pages), along with the expected adapter sequence. The random barcodes were then extracted from these identified adapter regions, and the adapter regions were trimmed from each read. Following trimming, the right end read was ~24bp, and the left end read was ~12bp.The trimmed reads were aligned to the H37Rv reference genome. The right end read was aligned by the BWA backtrack algorithm, using the “samse”, command. The left end read was aligned by a custom implementation of smith-waterman to locations within 500 bp of the right end read alignment. Reads in which at least one end failed to align, or in which the only alignment was disconcordant were discarded. Reads whose right end start could only be aligned more than 4bp away from a known enzyme restriction site were also discarded. The random barcode (NNANNANNN) were employed to control for artificially high counts generated by biased PCR amplification. Sequences with identical fusion positions and the same random barcode were counted as one unique sequence. Reads whose right end start was

aligned between 1 and 4 bp from a known enzyme restriction site were assumed to have been trimmed or aligned incorrectly, and were assigned to the known restriction enzyme site. The gene containing the alignment position of the right end start of each read was identified using the H37Rv RefSeq genome annotation released January 9 2012. Approximately 1% of all alignments landed in positions contained by more than one annotated gene, and the gene was assigned to the first gene in numerical order. Unambiguous unique sequences were each counted as a value of 1. Sequences that were ambiguous (could align to more than one location in the genome) were assigned to each of those fusion positions at a value of 1/# of potential alignment positions (e.g. for a sequence that could align to two potential fusion sites, each fusion was assigned a ½ read count). The Illumina HiSeq generated a different total number of reads per sample (average $1.5x10^7$ reads). All read counts were normalized to the total number of sequenced reads in that sample.

Statistical analysis used to identify in vivo exported proteins (spleen) Unique reads for each fusion site in the genome were counted in the output from the spleen of β-lactam treated mice, and the most abundant fusion site within each annotated gene was identified for each replicate. The lower abundance value between replicates was used as the representative abundance value for the gene in the subsequent statistical analysis to increase stringency and require that any identified gene was highly abundant in both samples. Log_{10} values were used to generate a histogram, which was bi-modal. A Gaussian mixture model was used to identify the mean and variance for each population, and determine the probability that each value identified with the higher or lower abundance population. Genes with a higher probability of belonging to the higher abundance group were identified as *in vivo* exported proteins. Three genes in regions with

enriched out of frame fusions were excluded from the results, resulting in 593 genes identified as encoding *in vivo* exported proteins.

Statistical analysis used to identify proteins exported significantly more in vivo than in vitro To identify *M. tuberculosis* clones that were β-lactam resistant both *in vivo* and *in vitro,* bacteria recovered from β-lactam treated mice were plated in parallel on agar media containing or lacking β-lactam antibiotics. From each *in vitro* plating condition, plasmids were isolated, sequenced, and unique reads for each fusion site counted, as described above. The most abundant fusion site within each annotated gene was identified for each condition and replicate. For *in vivo* β-lactam treated samples the lower abundance value between replicates was used as a representative abundance value for the gene, to identify the most stringent list of proteins exported in both experiments, as described above. For dual *in vivo* and *in vitro* β-lactam treated (recovered on β-lactam containing agar media *in vitro*) samples the higher abundance value between replicates was used as a representative abundance value for the gene, to identify the most stringent list of proteins that were not exported *in vitro* in either experiment. Pseudo-counts of 100 were added to the *in vivo* and *in vitro* β-lactam treated dataset. The log10 value of the ratio between *in vivo* treatment and *in vivo* plus *in vitro* treatment was calculated. The top and bottom 5% were trimmed for robustness. This data fit a normal uni-modal distribution, where genes of interest fell on the right shoulder of the curve (with high ratios of *in vivo* vs. *in vivo* and *in vitro* reads). A normal fit distribution was used to identify outliers, with higher ratios than would be predicted by chance. The Benjamini-Hochberg procedure was used to correct for multiple comparisons, and identified 41 genes with $p<0.0005$ (False Discovery Rate <0.05). Three genes in regions with enriched out of frame fusions were excluded from the results, resulting in 38 genes identified to encode for proteins exported significantly more *in vivo* than *in vitro*.

Analysis used to identify in vivo exported proteins (lungs). The lower starting burden in the lungs after intravenous infection resulted in incomplete coverage of the EXIT library in lungs (APPENDIX I) and prevented us from applying statistical modeling to the data as was done with the spleens. Unique reads for each fusion site in the genome were counted in the output from the lungs of β-lactam treated mice, and the most abundant fusion site within each annotated gene was identified separately for each replicate. The lower abundance value between replicates was used as a representative abundance value for the gene. Genes that displayed 3.5 fold more unique reads in the output than the input (3.5 fold enrichment) were used to predict proteins exported *in vivo* in the lungs. 3.5 fold enrichment correlated well with the enrichment level identified for the 593 statistically identified genes encoding proteins exported in the spleen.

Statistical analysis used to identify individual enriched fusion junctions for topology determination. The number of sequenced reads for each fusion site in the output from β-lactam treated mice was divided by the number of sequenced reads in the corresponding input for each experiment, with pseudo-counts of 10 added across the board. Values of zero were excluded from analysis. Log10 enrichment values were used to generate histograms, which produced a uni-modal distribution with a right shoulder of enriched sites. A Gaussian mixture model was fitted to the distribution using Mclust in R (Fraley *et al.*). The resulting mixture models had two peaks, one representing the majority of the sites, and a second, smaller peak representing points in the right shoulder. The derived cutoffs for Experiment A corresponded to an enrichment of log10 1.756 fold above input, and Experiment B corresponded to an enrichment of log10 3.203 fold. 2,516 individual fusion sites that were identified as being in the statistically enriched populations in both experiments were considered to be exported.

Construction of strains expressing HA tagged proteins for in vitro export analysis. Detailed plasmid information can be found on Table 2.1 and primer sequences are provided on Table 2.3. The *rv3707c* gene was PCR amplified using primers *rv3707c*_HA_F2 and *rv3707c*_HA_R1, the *rv1728c* gene with primers *rv1728c*_HA_F1 and *rv1728c*_HA_R1, and the *rv3811* gene with preimers *rv3811*_HA_F1 and *rv3811*_HA_R1, and PCR fragments were cloned into pCR2.1 (Invitrogen). The resulting plasmids were sequenced and confirmed error-free. Plasmids containing *rv3707c*, *rv1728c*, and *rv3811* were digested from pCR2.1 with XmnI and HindIII, gel purified, and ligated into MscI and HindIII digested pJSC77 (Glickman *et al.*, 2000), containing an in-frame C-terminal HA tag. Plasmids were transformed into *M. tuberculosis* H37Rv as previously described (Braunstein *et al.*, 2002).

Subcellular fractionation and Western blotting. M. tuberculosis cells were pelleted by centrifugation (1,900 x *g*), sterilized by irradiation (JL Shephard Mark I 137Cs irradiator, Department of Radiobiology, University of North Carolina at Chapel Hill), and sterilized *M. tuberculosis* cells were removed from BSL-3 containment. Subcellular fractionation was performed as previously described (Gibbons *et al.*, 2007). Cells were lysed by passage through a French pressure cell. Clarified whole cell lysates (WCL) were generated by centrifugation of lysed cellular material (1,900 x *g*) and then fractionated with differential ultracentrifugation. 27,000 x *g* for 30 minutes for cell wall (CW), 100,000 x *g* for 2 hours for the membrane (MEM), and the remaining soluble (SOL) fraction containing the cytoplasm. Fractions from equivalent original cell material were separated by SDS-PAGE and transferred to nitrocellulose membranes. Proteins were detected using the primary anti-HA antibody (1:25,000 from Covance) and secondary anti-mouse conjugated HRP (Biorad). HRP signal was detected using an enhanced chemiluminescence (Clarity Western ECL, Biorad).

Identification of export signals. Sequences were analyzed for transmembrane domains and signal peptides using TMHMM (Krogh *et al.*, 2001), Signal P (Petersen *et al.*, 2011), (Sutcliffe & Harrington, 2004), and (McDonough *et al.*, 2008). PE/PPE proteins were analyzed for YxxxD/E motifs as described in (Daleke *et al.*, 2012).

Attributions

E.F.P. and M.B. designed experiments and wrote the manuscript. E.F.P conducted all experiments except where noted below. D.W. provided initial insight into the project, constructed the library plasmid, pDW31, and assisted with construction of the library of plasmids in *E. coli*. J.D.H assisted in transformation of the EXIT library into *M. tuberculosis* and with murine experiments with the EXIT library. J.C.S. and T.R.I. carried out Illumina sequencing and T.R.I. performed statistical analysis of the EXIT dataset. D.O. and S.G. built the pipeline for analysis of the sequencing data.

Table 2.1 Plasmids used in this study			
Plasmid	**Antibiotic resistance**	**Notes**	**Source**
pMV261	kan	Multi-copy mycobacterial vector with *hsp60* promoter	(Stover *et al.*, 1991)
pJSC77	kan	Multi-copy mycobacterial vector, HA tag cloned into pMV261	(Glickman *et al.*, 2000)
pEP207	kan	P_{hsp60}-*rv3707c* cloned into pJSC77	This study
pEP210	kan	P_{hsp60}-*rv1728* cloned into pJSC77	This study
pEP213	kan	P_{hsp60}-*rv3811* cloned into pJSC77	This study
pMB219	kan	Multi-copy mycobacterial vector	(McCann *et al.*, 2007)
pYUB412	hyg	Single copy integrating mycobacterial vector	(Bange *et al.*, 1999)
pDW1	hyg	Multi-copy mycobacterial vector containing backbone from pMB219 and hygromycin resistance cassette from pYUB412	This study
pDW10	hyg	pDW1 with extraneous sequence deleted	This study
pJES110	kan, amp	*'blaTEM* reporter in cloning vector pCR2.1 Topo (Invitrogen)	This study
pDW14	hyg	Multi-copy mycobacterial vector, with *'blaTEM* reporter downstream of ClaI restriction site	This study
pDW31	hyg	Multi-copy mycobacterial vector, with *'blaTEM* reporter downstream of ClaI restriction site	This study

Table 2.1. Plasmids used in this study.

Table 2.2 Bacterial strains used in this study		
***M. tuberculosis* strains**	**Description**	**Source**
PM638	*M. tuberculosis* H37Rv *ΔblaC*	(Flores *et al.*, 2005)
EXIT Library	PM638 + pDW31 containing PM638 genomic DNA fragments (~500-5,000 bp) in ClaI site. Strain is hyg^R	This study
MBTB542	H37Rv + pEP207 (expressing *rv3707c-HA*). Strain is kan^R	This study
MBTB543	H37Rv + pEP210 (expressing *rv1728-HA*). Strain is kan^R	This study
MBTB544	H37Rv + pEP213 (expressing *rv3811-HA*). Strain is kan^R	This study

Table 2.2. Bacterial strains used in this study.

Table 2.3 Primers used in this study		
Name	**Primer sequence**	**Source**
Adaptor 1.1	TACCACGACCA-NH2	(Long *et al.*, 2015)
Adaptor 2.1	ATGATGGCCGGTGGATTTGTGNNANNANNNTGGTCGTGGTAT	(Long *et al.*, 2015)
'Bla IL	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTACTTTCACCAGCGTTTC	This study
'Bla IL 3b	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTTACTTTCACCAGCGTTTC	This study
'Bla IL 4b	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGATATTTACTTTCACCAGCGTTTC	This study
'Bla IL 5b	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATCTATTTACTTTCACCAGCGTTTC	This study
Adaptor 1	CAAGCAGAAGACGGCATACGAGATAAGTAGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATGATGGCCGGTGGATTTGTG	(Long *et al.*, 2015)
Adaptor 2	CAAGCAGAAGACGGCATACGAGATACACGATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATGATGGCCGGTGGATTTGTG	(Long *et al.*, 2015)
Adaptor 3	CAAGCAGAAGACGGCATACGAGATCGCGCGGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATGATGGCCGGTGGATTTGTG	(Long *et al.*, 2015)
*rv3707c*_HA_F2	GAATGCCTTCCTGCGAATCGGTCCGACGGCCGGTAC	This study
*rv3707c*_HA_R1	AAGCTTGCGGGTCTGACCAGGGCTTGAAC	This study
*rv1728c*_HA_F1	CGAATGCCTTCATGAGCGTGAACGGCTTGC	This study
*rv1728c*_HA_R1	AAGCTTGTTCTGGCGGCGTAGGGCTC	This study
*rv3811*_HA_F1	GAATGCCTTCGCAGCGACCGTCGTCATCGTCGCGTGGATAG	This study
*rv3811*_HA_R1	AAGCTTGGGTGATCGGATGCGTTGGCAGCGTGAAG	This study

Table 2.3. Primers used in this study. Primers used to prepare fragments for Illumina sequencing are color coded for additional information. The adaptor ligation sites represent where the double stranded adaptor primers attach for Adaptor 1.1 and Adaptor 2.1 (pink), or where the PCR amplification primer has homology to the adaptor for PCR amplification (yellow, Adaptor 2.1 and Adaptor primers). Two barcodes were used, a random barcode integrated into the Adaptor 2.1 sequence (bright green) to be able to quantify PCR biases in amplification, and an index or multiplexing barcode to allow for multiple sequences to be simultaneously sequenced in Adaptor 1, 2, and 3 primers (purple). PCR amplification using 'Bla IL and Adaptor 1,2,3 primers was used to attach the sequences necessary for attachment to the Illumina chip (blue) and for sequencing with the Illumina primers (red, Read 1 sequence in Bla IL primers Read 2 sequence in Adaptor primers). The 'Bla IL primers additionally possessed homology to the 'BlaTEM reporter for PCR amplification (dark green), and a variable site for Illumina sequencing (grey).

Table 2.4 Proof of principle enrichment studies			
	Initial % β-lactam^R	**% β-lactam^R after infection with no treatment**	**% β-lactam^R after infection with *in vivo* β-lactam treatment**
99:1 'BlaTEM: sp-'BlaTEM	1%	1%	69%
EXIT Library	1%	1%	76%

Table 2.4. Proof of principle enrichment studies. Mice were infected by tail vein injection with a mixture of *M. tuberculosis* strains producing a non-exported 'BlaTEM reporter ('BlaTEM, black), or an exported signal peptide fused 'BlaTEM reporter (sp-'BlaTEM, red) or in a 99:1 ratio ('BlaTEM:sp-'BlaTEM). The inoculum was plated in parallel on normal agar media and agar media containing β-lactam antibiotics, to determine the initial β-lactam resistance frequency. Half of the mice were treated with the β-lactam antibiotic amoxicillin and a synergistic drug probenecid twice daily by oral gavage, while half remained untreated. On day 14 mice were sacrificed and spleens were homogenized and plated again in parallel on normal media and agar media containing β-lactam antibiotics, to determine the final β-lactam resistance frequency. This experiment was repeated with the EXIT library to determine initial β-lactam resistance frequency as well as β-lactam resistance frequency following *in vivo* β-lactam resistance treatment.

Table 2.5 Known Exported Proteins (Adapted from McCann *et al.* 2011)			
Genome Designation	**gene**	**standard_name**	**function**
Rv0064			UNKNOWN
Rv0072			THOUGHT TO BE INVOLVED IN ACTIVE TRANSPORT OF GLUTAMINE ACROSS THE MEMBRANE (IMPORT). RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv0092	ctpA		CATION-TRANSPORTING ATPASE; POSSIBLY CATALYZES THE TRANSPORT OF A CATION (POSSIBLY COPPER) WITH THE HYDROLYSE OF ATP [CATALYTIC ACTIVITY: ATP + H(2)O + CATION(IN) = ADP + PHOSPHATE + CATION(OUT)].
Rv0116c			UNKNOWN
Rv0125	pepA	mtb32a	UNKNOWN; POSSIBLY HYDROLYZES PEPTIDES AND/OR PROTEINS (SEEMS TO CLEAVE PREFERENTIALLY AFTER SERINE RESIDUES).
Rv0169	mce1A	mce1	UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST CELL INVASION (ENTRY AND SURVIVAL INSIDE MACROPHAGES).
Rv0170	mce1B	mceD	UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST CELL INVASION.
Rv0171	mce1C		UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST CELL INVASION.
Rv0172	mce1D		UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST CELL INVASION.
Rv0173	lprK	mce1E	UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST CELL INVASION.
Rv0174	mce1F		UNKNOWN, BUT THOUGHT INVOLVED IN HOST CELL INVASION.
Rv0175			UNKNOWN
Rv0178			UNKNOWN
Rv0199			UNKNOWN
Rv0265c		fecB2	THOUGHT TO BE INVOLVED IN IRON TRANSPORT ACROSS THE MEMBRANE (IMPORT).
Rv0312			UNKNOWN
Rv0346c	ansP2	aroP2	THOUGHT TO BE INVOLVED IN TRANSPORT OF L-ASPARAGINE ACROSS THE MEMBRANE. RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv0361			UNKNOWN
Rv0402c	mmpL1		UNKNOWN. THOUGHT TO BE INVOLVED IN FATTY ACID TRANSPORT.
Rv0412c			UNKNOWN
Rv0418	lpqL		UNKNOWN; HYDROLYZES PEPTIDES AND/OR PROTEINS.
Rv0432	sodC		DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS [CATALYTIC ACTIVITY: 2

			superoxide + 2 H+ = O2 + H2O2].
Rv0450c	mmpL4		UNKNOWN. THOUGHT TO BE INVOLVED IN FATTY ACID TRANSPORT.
Rv0453	PPE11		UNKNOWN
Rv0483	lprQ		UNKNOWN
Rv0506	mmpS2		UNKNOWN
Rv0517			UNKNOWN; PROBABLY INVOLVED IN CELLULAR METABOLISM.
Rv0583c	lpqN		UNKNOWN
Rv0594	mce2F		UNKNOWN, BUT THOUGHT INVOLVED IN HOST CELL INVASION.
Rv0676c	mmpL5		UNKNOWN. THOUGHT TO BE INVOLVED IN FATTY ACID TRANSPORT.
Rv0677c	mmpS5		UNKNOWN
Rv0783c	emrB		TRANSLOCASE THAT CONFERS RESISTANCE TO SUBSTANCES OF HIGH HYDROPHOBICITY. INVOLVED IN TRANSPORT OF MULTIDRUG ACROSS THE MEMBRANE (EXPORT): MULTIDRUG RESISTANCE BY AN EXPORT MECHANISM. RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv0837c			UNKNOWN
Rv0876c			UNKNOWN
Rv0917	betP		HIGH-AFFINITY UPTAKE OF GLYCINE BETAINE. SUPPOSED RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv0931c	pknD	mbk	INVOLVED IN SIGNAL TRANSDUCTION (VIA PHOSPHORYLATION). THOUGHT TO REGULATE PHOSPHATE TRANSPORT. CAN PHOSPHORYLATE THE PEPTIDE SUBSTRATE MYELIN BASIC PROTEIN (MBP) AT SERINE AND THREONINE RESIDUES. CAN BE AUTOPHOSPHORYLATED ON THREONINE RESIDUES [CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN].
Rv0934	pstS1	phoS1; phoS	INVOLVED IN ACTIVE TRANSPORT OF INORGANIC PHOSPHATE ACROSS THE MEMBRANE (IMPORT). THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE TRANSPORT.
Rv0983	pepD	mtb32b	UNKNOWN; POSSIBLY HYDROLYZES PEPTIDES AND/OR PROTEINS (SEEMS TO CLEAVE PREFERENTIALLY AFTER SERINE RESIDUES).
Rv1004c			UNKNOWN
Rv1009	rpfB		THOUGHT TO PROMOTE THE RESUSCITATION AND GROWTH OF DORMANT, NONGROWING CELL. COULD ALSO STIMULATES THE GROWTH OF SEVERAL OTHER HIGH G+C GRAM+ ORGANISMS, e.g. Mycobacterium avium, Mycobacterium bovis (BCG), Mycobacterium kansasii, Mycobacterium smegmatis.

Rv1078	pra		UNKNOWN
Rv1096			PROBABLY INVOLVED IN CARBOHYDRATE DEGRADATION. May Hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety.
Rv1157c			UNKNOWN
Rv1164	narI		NITRATE REDUCTION [CATALYTIC ACTIVITY: Nitrite + acceptor = nitrate + reduced acceptor].
Rv1174c	TB8.4		UNKNOWN FUNCTION (SECRETED PROTEIN)
Rv1230c			UNKNOWN
Rv1239c	corA		THOUGHT TO BE INVOLVED IN TRANSPORT OF MAGNESIUM AND COBALT IONS ACROSS THE MEMBRANE. RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv1319c			THOUGHT TO PLAY AN ESSENTIAL ROLES IN REGULATION OF CELLULAR METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER, CAMP [CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE].
Rv1368	lprF		UNKNOWN
Rv1419			UNKNOWN
Rv1435c			UNKNOWN
Rv1567c			UNKNOWN
Rv1591			UNKNOWN
Rv1635c			UNKNOWN
Rv1648			UNKNOWN
Rv1707			UNKNOWN; POSSIBLY INVOLVED IN TRANSPORT OF SULFATE ACROSS THE MEMBRANE.
Rv1728c			UNKNOWN
Rv1743	pknE		INVOLVED IN SIGNAL TRANSDUCTION (VIA PHOSPHORYLATION). THOUGHT TO BE INVOLVED IN MEMBRANE TRANSPORT [CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN].
Rv1779c			UNKNOWN
Rv1819c			THOUGHT TO BE INVOLVED IN ACTIVE TRANSPORT OF DRUGS ACROSS THE MEMBRANE (EXPORT): MULTIDRUGS RESISTANCE BY AN EXPORT MECHANISM. RESPONSIBLE FOR ENERGY COUPLING TO THE TRANSPORT SYSTEM AND FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv1836c			UNKNOWN
Rv1860	apa	mpt32; modD	UNKNOWN (COULD MEDIATE BACTERIAL ATTACHMENT TO HOST CELLS).

Rv1886c	fbpB	mpt59; 85B	INVOLVED IN CELL WALL MYCOLOYLATION. PROTEINS OF THE ANTIGEN 85 COMPLEX ARE RESPONSIBLE FOR THE HIGH AFFINITY OF MYCOBACTERIA TO FIBRONECTIN. POSSESSES A MYCOLYLTRANSFERASE ACTIVITY REQUIRED FOR THE BIOGENESIS OF TREHALOSE DIMYCOLATE (CORD FACTOR), A DOMINANT STRUCTURE NECESSARY FOR MAINTAINING CELL WALL INTEGRITY.
Rv1887			UNKNOWN
Rv1891			UNKNOWN
Rv1984c	cfp21		HYDROLYZES CUTIN.
Rv2040c			Thought to be involved in active transport of sugar across the membrane (import). Responsible for the translocation of the substrate across the membrane.
Rv2068c	blaC		hydrolyses beta-lactams to generate corresponding beta-amino acid [CATALYTIC ACTIVITY: A BETA-LACTAM + H(2)O = A SUBSTITUTED BETA-AMINO ACID].
Rv2080	lppJ		UNKNOWN
Rv2113			UNKNOWN
Rv2127	ansP1		Involved in L-asparagine transport.
Rv2200c	ctaC		INVOLVED IN AEROBIC RESPIRATION. SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA HEME A AND CU(A) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND CU(B) (BY SIMILARITY).
Rv2203			UNKNOWN
Rv2240c			UNKNOWN
Rv2264c			UNKNOWN
Rv2284	lipM		Hydrolysis of lipids (bound ester).
Rv2290	lppO		UNKNOWN
Rv2301	cut2	cfp25	HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE).
Rv2320c	rocE		THOUGHT TO BE INVOLVED IN TRANSPORT OF CATIONIC AMINO ACID (ESPECIALLY ARGININE AND ORNITHINE) ACROSS THE MEMBRANE. RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv2339	mmpL9		UNKNOWN. THOUGHT TO BE INVOLVED IN FATTY ACID TRANSPORT.
Rv2394	ggtB		PLAYS A KEY ROLE IN THE GAMMA-GLUTAMYL CYCLE, A PATHWAY FOR THE SYNTHESIS AND DEGRADATION OF GLUTATHIONE [CATALYTIC ACTIVITY: 5-L-GLUTAMYL)-PEPTIDE + AN AMINO ACID = PEPTIDE + 5-L-GLUTAMYL-AMINO ACID].

Rv2443	dctA		INVOLVED IN THE TRANSPORT OF DICARBOXYLATES SUCH AS SUCCINATE, FUMARATE, AND MALATE FROM THE PERIPLASM ACROSS THE INNER MEMBRANE. RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv2585c			UNKNOWN
Rv2599			UNKNOWN
Rv2639c			UNKNOWN
Rv2721c			UNKNOWN
Rv2905	lppW		UNKNOWN
Rv2938	drrC		PROBABLY INVOLVED IN ACTIVE TRANSPORT OF ANTIBIOTIC AND PHTHIOCEROL DIMYCOCEROSATE (DIM) ACROSS THE MEMBRANE (EXPORT). DRRA|Rv2934|MTCY19H9.04, DRRB|Rv2937|MTCY19H9.05 AND DRRC MAY ACT JOINTLY TO CONFER DAUNORUBICIN AND DOXORUBICIN RESISTANCE BY AN EXPORT MECHANISM. PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE AND LOCALIZATION OF DIM INTO THE CELL WALL.
Rv2945c	lppX		UNKNOWN
Rv3036c	TB22.2		UNKNOWN
Rv3090			UNKNOWN
Rv3103c			UNKNOWN
Rv3209			UNKNOWN
Rv3253c			THOUGHT TO BE INVOLVED IN CATIONIC AMINO ACID TRANSPORT ACROSS THE MEMBRANE. RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv3267			UNKNOWN
Rv3390	lpqD		UNKNOWN
Rv3413c			UNKNOWN
Rv3476c	kgtP		INVOLVED IN ACTIVE TRANSPORT OF DICARBOXYLIC ACID ACROSS THE MEMBRANE. RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
Rv3484	cpsA		NOT KNOW.
Rv3494c	mce4F		UNKNOWN, BUT THOUGHT INVOLVED IN HOST CELL INVASION.
Rv3496c	mce4D		UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST CELL INVASION.
Rv3497c	mce4C		UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST CELL INVASION.
Rv3498c	mce4B		UNKNOWN, BUT THOUGHT TO BE INVOLVED IN HOST CELL INVASION.
Rv3584	lpqE		UNKNOWN
Rv3689			UNKNOWN

Rv3779			UNKNOWN
Rv3802c			UNKNOWN
Rv3835			UNKNOWN
Rv3869			UNKNOWN
Rv3901c			UNKNOWN
Rv3910			UNKNOWN

Table 2.5. Known exported proteins. 111 *M. tuberculosis* proteins were identified in (McCann *et al.*, 2011)to be exported *in vitro*. These proteins were used as known positives, or known exported proteins, to compare to EXIT identified exported proteins.

Table 2.6 50 Most abundant cytoplasmic proteins (adapted from PaxDB (Wang *et al.*, 2015; Wang *et al.*, 2012))				
Abundance rank	**Rv number**	**Name**	**Estimated PPM**	**Function**
1	Rv2031c	acr	125706	heat shock protein HspX
2	Rv3418c	Rv3418c	96990	co-chaperonin GroES; Binds to Cpn60 in the presence of Mg-ATP and suppresses the ATPase activity of the latter
3	Rv0440	Rv0440	37230	chaperonin GroEL; Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity)
6	Rv0652	rplL	11450	50S ribosomal protein L7/L12; Seems to be the binding site for several of the factors involved in protein synthesis and appears to be essential for accurate translation (By similarity)
7	Rv1470	Rv1470	10999	thioredoxin TrxC (TRX) (MPT46); Participates in various redox reactions through the reversible oxidation of its active center dithiol to a disulfide and catalyzes dithiol-disulfide exchange reactions (By similarity)
8	Rv2244	acpM	10397	acyl carrier protein; Acyl carrier protein involved in meromycolate extension (By similarity)
9	Rv0350	dnaK	9967	molecular chaperone DnaK; Acts as a chaperone (By similarity)
11	Rv0685	tuf	7356	elongation factor Tu; This protein promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis
14	Rv3648c	cspA	6541	cold shock protein A
19	Rv3028c	etfA	4429	electron transfer flavoprotein subunit alpha; The electron transfer flavoprotein serves as a specific electron acceptor for other dehydrogenases. It transfers the electrons to the main respiratory chain via ETF-ubiquinone oxidoreductase (ETF dehydrogenase) (By similarity)
20	Rv1133c	metE	4164	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase; Catalyzes the transfer of a methyl group from 5-methyltetrahydrofolate to homocysteine resulting in methionine formation (By similarity)
22	Rv2623	TB31.7	3757	hypothetical protein
23	Rv2220	glnA1	3713	glutamine synthetase GLNA1 (glutamine synthase) (GS-I)
25	Rv2145c	ag84	3535	hypothetical protein
26	Rv1872c	lldD2	3477	L-lactate dehydrogenase (cytochrome) LldD2
29	Rv3118	sseC1	3200	hypothetical protein
30	Rv2744c	Rv2744c	3194	hypothetical protein

31	Rv0475	hbhA	3091	iron-regulated heparin binding hemagglutinin HbhA (adhesin); Required for extrapulmonary dissemination. Mediates adherence to epithelial cells by binding to sulfated glycoconjugates present at the surface of these cells; binds heparin, dextran sulfate, fucoidan and chondroitin sulfate. Promotes hemagglutination of erythrocytes of certain host species. Induces mycobacterial aggregation
33	Rv2986c	Rv2986c	2920	DNA-binding protein HU; Histone-like DNA-binding protein which is capable of wrapping DNA to stabilize it, and thus to prevent its denaturation under extreme environmental conditions (By similarity)
34	Rv1827	Rv1827	2870	hypothetical protein
35	Rv1498A	Rv1498A	2851	hypothetical protein
36	Rv3248c	Rv3248c	2796	S-adenosyl-L-homocysteine hydrolase
37	Rv3846	Rv3846	2772	superoxide dismutase [Fe] SODA; Destroys radicals which are normally produced within the cells and which are toxic to biological systems
40	Rv1392	metK	2683	S-adenosylmethionine synthetase; Catalyzes the formation of S-adenosylmethionine from methionine and ATP. The overall synthetic reaction is composed of two sequential steps, AdoMet formation and the subsequent tripolyphosphate hydrolysis which occurs prior to release of AdoMet from the enzyme (By similarity)
41	Rv1738	Rv1738	2671	hypothetical protein
42	Rv3841	bfrB	2649	bacterioferritin BfrB
44	Rv0379	secE2	2569	protein transport protein
45	Rv3716c	Rv3716c	2551	hypothetical protein
46	Rv2159c	Rv2159c	2540	hypothetical protein
47	Rv1932	tpx	2519	thiol peroxidase; Has antioxidant activity. Could remove peroxides or H(2)O(2) (By similarity)
48	Rv1636	TB15.3	2469	hypothetical protein
49	Rv3457c	rpoA	2456	DNA-directed RNA polymerase subunit alpha; DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates
50	Rv3029c	etfB	2426	electron transfer flavoprotein subunit beta; The electron transfer flavoprotein serves as a specific electron acceptor for other dehydrogenases. It transfers the electrons to the main respiratory chain via ETF-ubiquinone oxidoreductase (ETF dehydrogenase) (By similarity)
52	Rv1310	atpD	2158	F0F1 ATP synthase subunit beta; Produces ATP from ADP in the presence of a proton gradient across the membrane. The catalytic sites are hosted primarily by the beta subunits (By

				similarity)
53	Rv1177	fdxC	2140	ferredoxin FdxC
54	Rv2461c	clpP1	2124	ATP-dependent Clp protease proteolytic subunit; Cleaves peptides in various proteins in a process that requires ATP hydrolysis. Has a chymotrypsin-like activity. Plays a major role in the degradation of misfolded proteins (By similarity)
55	Rv0009	ppiA	2116	iron-regulated peptidyl-prolyl cis-trans isomerase A; PPIases accelerate the folding of proteins. It catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides (By similarity)
56	Rv2111c	Rv2111c	2109	hypothetical protein; Protein modifier that is covalently attached to lysine residues of substrate proteins, thereby targeting them for proteasomal degradation. The tagging system is termed pupylation. Identified substrates are the fabD, panB and mpa proteins
57	Rv2215	sucB	2095	dihydrolipoamide acetyltransferase; The 2-oxoglutarate dehydrogenase complex catalyzes the overall conversion of 2-oxoglutarate to succinyl-CoA and CO(2). It contains multiple copies of three enzymatic components: 2-oxoglutarate dehydrogenase (E1), dihydrolipoamide succinyltransferase (E2) and lipoamide dehydrogenase (E3) (By similarity)
58	Rv3458c	rpsD	2094	30S ribosomal protein S4; One of the primary rRNA binding proteins, it binds directly to 16S rRNA where it nucleates assembly of the body of the 30S subunit (By similarity)
59	Rv0569	Rv0569	2071	hypothetical protein
60	Rv1308	atpA	2064	F0F1 ATP synthase subunit alpha; Produces ATP from ADP in the presence of a proton gradient across the membrane. The alpha chain is a regulatory subunit
61	Rv1436	gap	2035	glyceraldehyde-3-phosphate dehydrogenase
62	Rv3417c	Rv3417c	1978	chaperonin GroEL; Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity)
63	Rv1388	mihF	1970	putative integration host factor MIHF
64	Rv0054	ssb	1942	single-stranded DNA-binding protein; This protein is essential for replication of the chromosome. It is also involved in DNA recombination and repair (By similarity)

67	Rv0683	rpsG	1922	30S ribosomal protein S7; One of the primary rRNA binding proteins, it binds directly to 16S rRNA where it nucleates assembly of the head domain of the 30S subunit. Is located at the subunit interface close to the decoding center, probably blocks exit of the E-site tRNA (By similarity)
68	Rv1211	Rv1211	1910	hypothetical protein
71	Rv0363c	fba	1877	fructose-bisphosphate aldolase; Catalyzes the aldol condensation of dihydroxyacetone phosphate (DHAP or glycerone-phosphate) with glyceraldehyde 3- phosphate (G3P) to form fructose 1,6-bisphosphate (FBP) in gluconeogenesis and the reverse reaction in glycolysis (By similarity)
73	Rv3628	ppa	1860	inorganic pyrophosphatase

Table 2.6. Most abundant cytoplasmic proteins. The top 50 most abundant cytoplasmic proteins were identified by PaxDB, a database calculating the relative abundance of proteins from published proteomics datasets (Wang *et al.*, 2015; Wang *et al.*, 2012). Proteins identified as abundant in PaxDB but containing predicted export signals (transmembrane domains and signal peptides were predicted by SignalP, TatP, TMHMM, (Sutcliffe & Harrington, 2004), and (McDonough *et al.*, 2008)), or proteins known to be exported (e.g. EsxA and EsxB) were excluded, to generate a list of the top 50 most abundant cytoplasmic proteins in *M. tuberculosis*. These abundant cytoplasmic proteins were used as known negatives, or known cytoplasmic proteins, to compare to EXIT identified proteins.

Table 2.7 EXIT identified proteins with no *in silico* export signals			
Gene number	**Name**	**Product**	**MS**
Rv0518		POSSIBLE EXPORTED PROTEIN	CF 2, CF 3, MEM 8
Rv0787		HYPOTHETICAL PROTEIN	CF 2, CF 3, MEM 7, MEM 9
Rv0822c		CONSERVED HYPOTHETICAL PROTEIN	CF 2, CW 14
Rv0907		CONSERVED HYPOTHETICAL PROTEIN	CF 2, CF 3, MEM 5, MEM 8, MEM 9, CW 14
Rv0950c		CONSERVED HYPOTHETICAL PROTEIN	
Rv1026		CONSERVED HYPOTHETICAL PROTEIN	
Rv1728c		CONSERVED HYPOTHETICAL PROTEIN	
Rv1823		CONSERVED HYPOTHETICAL PROTEIN	MEM 8, MEM 9, CW 14
Rv1832	gcvB	Probable glycine dehydrogenase gcvB (Glycine decarboxylase) (Glycine cleavage system P-protein)	CF 2, MEM 4, MEM 8, MEM 9, WCL 15, SOL 16
Rv1887		HYPOTHETICAL PROTEIN	CF 3, MEM 7
Rv1891		CONSERVED HYPOTHETICAL PROTEIN	CF 2, CF 3, MEM 7, MEM 8
Rv2088	pknJ	PROBABLE TRANSMEMBRANE SERINE/THREONINE-PROTEIN KINASE J PKNJ (PROTEIN KINASE J) (STPK J)	MEM 8, MEM 9, CW 14
Rv2240c		HYPOTHETICAL PROTEIN	CF 2, CF 3, MEM 8, MEM 9, CW 14
Rv2264c		conserved hypothetical proline rich protein	CW 14
Rv2300c		CONSERVED HYPOTHETICAL PROTEIN	MEM 8, MEM 9
Rv2380c	mbtE	PEPTIDE SYNTHETASE MBTE (PEPTIDE SYNTHASE)	MEM 4, MEM 9, CW 13, SOL 16
Rv3035		CONSERVED HYPOTHETICAL PROTEIN	MEM 8, MEM 9, CW 14
Rv3067		CONSERVED HYPOTHETICAL PROTEIN	CF 2
Rv3123		HYPOTHETICAL PROTEIN	MEM 9

Rv3274c	fadE25	PROBABLE ACYL-CoA DEHYDROGENASE FADE25	CF 2, CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16
Rv3343c	PPE54	PPE FAMILY PROTEIN	
Rv3350c	PPE56	PPE FAMILY PROTEIN	
Rv3478	PPE60, mtb39c	PE FAMILY PROTEIN	MEM 8, MEM 9, CW 13, CW 14
Rv3526		POSSIBLE OXIDOREDUCTASE	
Rv3596c	clpC1, clpC	PROBABLE ATP-DEPENDENT PROTEASE ATP-BINDING SUBUNIT CLPC1	CF 3, MEM 4, MEM 6, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16
Rv3654c		CONSERVED HYPOTHETICAL PROTEIN	
Rv3667	acs	ACETYL-COENZYME A SYNTHETASE ACS (ACETATE--CoA LIGASE) (ACETYL-CoA SYNTHETASE) (ACETYL-CoA SYNTHASE) (ACYL-ACTIVATING ENZYME) (ACETATE THIOKINASE) (ACETYL-ACTIVATING ENZYME) (ACETATE--COENZYME A LIGASE) (ACETYL-COENZYME A SYNTHASE)	MEM 4, MEM 9, CW 13, CW 14, WCL 15, SOL 16
Rv3691		CONSERVED HYPOTHETICAL PROTEIN	MEM 4, MEM 8, MEM 9, CW 14
Rv3707c		CONSERVED HYPOTHETICAL PROTEIN	MEM 9
Rv3811	csp	CONSERVED HYPOTHETICAL PROTEIN	
Rv3822			CW 14
Rv3912		HYPOTHETICAL ALANINE RICH PROTEIN	MEM 9

Table 2.7. EXIT identified proteins with no *in silico* export signals. Proteins identified as exported by EXIT were analyzed for export signals by SignalP, TatP, TMHMM, (McDonough *et al.*, 2008), and (Sutcliffe & Harrington, 2004). 32 remained that contained no predicted export signal. Proteins are identified by their genome designation (column 1), name (column 2), and annotation from the H37Rv RefSeq genome annotation released January 9 2012. Column 4 identifies all previous *in vitro* MS based published methodologies which identified a given protein as exported, with the fractions in which the

protein was identified: culture filtrate (CF), membrane (MEM), cell wall (CW), soluble (SOL), or whole cell lysate (WCL). Mass spectrometry based approaches to identify exported proteins in fractions: CF: 1 (Rosenkrands *et al.*, 2000b), 2 (Malen *et al.*, 2011), 3 (Bell *et al.*, 2012). MEM: 4 (Gu *et al.*, 2003), 5 (Xiong *et al.*, 2005), 6 (Mawuenyega *et al.*, 2005), 7 (Malen *et al.*, 2007), 8 (Malen *et al.*, 2011), 9 (Gunawardena *et al.*, 2013), 10 (Bell *et al.*, 2012). CW: 11 (Mawuenyega *et al.*, 2005), 12 (Wolfe *et al.*, 2010), 13 (Bell *et al.*, 2012). WCL: 15 (Bell *et al.*, 2012). SOL: 16 (Bell *et al.*, 2012).

Table 2.8 EXIT exported fusions in unannotated regions			
Strand location	**Upstream annotated gene**	**Genome position**	**Distance in aa from paired site(s)**
-	Rv0066c	65092	-
-	Rv0066c	65050	14
+	Rv0397	476536	-
+	Rv0397	476632	32
-	Rv2307A	2578951	-
-	Rv2307A	2578942	3
-	Rv2307A	2578876	22
-	Rv2307c	2575499	-
-	Rv2307c	2575451	16
-	Rv2307c	2575403	16
-	Rv2307c	2575385	6
-	Rv2307c	2575379	2
-	Rv2307c	2575358	7
-	Rv2307c	2575340	6
+	Rv2964	3318177	-
+	Rv2964	3318183	2
+	Rv2964	3318189	2
+	Rv2964	3318198	3
+	Rv3033	3395236	
+	Rv3033	3395269	11
+	Rv3033	3395302	11

Table 2.8. EXIT exported fusions in unannotated regions. Enriched sites data located in intergenic non-annotated regions of the genome were identified. Regions of the genome likely to contain a non-annotated exported protein were identified as containing at least two statistically enriched sites in frame of each other at a maximum distance of 100 bp apart. Six intergenic regions were identified, labeled A-F, containing 21 statistically enriched sites. The genome location and upstream annotated gene are identified, as well as the amino acid distance to additional enriched sites identified in the same region.

Table 2.9 EXIT exported proteins only identified in the lungs			
Gene number	**Name**	**Product**	***in silico* Export Signal**
Rv0187		PROBABLE O-METHYLTRANSFERASE	
Rv0578c	PE_PGRS7	PE-PGRS FAMILY PROTEIN	SP, YxxxD/E
Rv1091	PE_PGRS22	PE-PGRS FAMILY PROTEIN	YxxxD/E
Rv1371		PROBABLE CONSERVED MEMBRANE PROTEIN	TM
Rv1818c	PE_PGRS33	PE-PGRS FAMILY PROTEIN	SP, YxxxD/E
Rv2196	qcrB	Probable Ubiquinol-cytochrome C reductase QcrB (cytochrome B subunit)	TM
Rv2490c	PE_PGRS43	PE-PGRS FAMILY PROTEIN	SP, YxxxD/E
Rv3101c	ftsX	PUTATIVE CELL DIVISION PROTEIN FTSX (SEPTATION COMPONENT-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER)	TM

Table 2.9. EXIT exported proteins only identified in the lungs. Abundance was determined individually for each gene based on sequenced read counts for the most highly abundant fusion site in each experiment for the lungs. Genes that displayed 3.5 fold higher abundance in the lungs after *in vivo* β-lactam treatment than the input in both experiments were identified as exported *in vivo* in the lungs. 282 proteins were identified as exported in the lungs. The eight proteins reported above were only identified as exported in the lungs and not identified as exported in the spleen.

Table 2.10 EXIT identified proteins exported significantly more *in vivo* than *in vitro*						
Gene number	**Name**	**Product**	***in silico***	**Essential**	**Up *in vivo***	**q value**
Rv0011c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	TM			0.000
Rv0261c	narK3	PROBABLE INTEGRAL MEMBRANE NITRITE EXTRUSION PROTEIN NARK3 (NITRITE FACILITATOR)	TM			0.003
Rv0490	senX3	PUTATIVE TWO COMPONENT SENSOR HISTIDINE KINASE SENX3	TM	Mouse 25		0.028
Rv0506	mmpS2	PROBABLE CONSERVED MEMBRANE PROTEIN MMPS2	SP, TM			0.007
Rv0559c		POSSIBLE CONSERVED SECRETED PROTEIN	SP			0.008
Rv0593	lprL, mce2E	POSSIBLE MCE-FAMILY LIPOPROTEIN LPRL (MCE-FAMILY LIPOPROTEIN MCE2E)	Lipo, TM			0.006
Rv0615		PROBABLE INTEGRAL MEMBRANE PROTEIN	TM			0.017
Rv0713		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	TM			0.001
Rv0817c		PROBABLE CONSERVED EXPORTED PROTEIN	SP, TM	in vitro 28, in vitro 29		0.005
Rv0846c		PROBABLE OXIDASE	SP, Tat SP, Lipo			0.000
Rv0892		PROBABLE MONOOXYGENASE	TM		Macrophage 31	0.008
Rv1026		CONSERVED HYPOTHETICAL PROTEIN		in vitro 28		0.022
Rv1145	mmpL13a	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL13A	TM			0.000
Rv1266c	pknH	PROBABLE TRANSMEMBRANE SERINE/THREONINE-PROTEIN KINASE H PKNH (PROTEIN KINASE H) (STPK H)	TM			0.029
Rv1508c		Probable membrane protein	TM		Mouse 34	0.007
Rv1517		CONSERVED HYPOTHETICAL TRANSMEMBRANE PROTEIN	Tat SP, TM			0.015
Rv1639c		CONSERVED HYPOTHETICAL MEMBRANE PROTEIN	TM			0.007

Rv1737c	narK2	POSSIBLE NITRATE/NITRITE TRANSPORTER NARK2	TM	Mouse 27	Macrophage 31	0.000
Rv1739c		PROBABLE SULPHATE-TRANSPORT TRANSMEMBRANE PROTEIN ABC TRANSPORTER	TM		Macrophage 31	0.043
Rv1965	yrbE3B	CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBE3B	TM	Macaque 22	Macrophage 31	0.001
Rv1969	mce3D	MCE-FAMILY PROTEIN MCE3D	SP, TM			0.006
Rv2138	lppL	Probable conserved lipoprotein LppL	SP, Lipo, TM	Mouse 27, in vitro 28, in vitro 29		0.006
Rv2144c		Probable transmembrane protein	TM		Mouse 34	0.001
Rv2273		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	TM		Macrophage 30	0.000
Rv2284	lipM	Probable esterase LipM	TM		Mouse 37	0.024
Rv2330c	lppP	PROBABLE LIPOPROTEIN LPPP	SP, Lipo, TM	Macrophage 23	Macrophage 31	0.007
Rv2380c	mbtE	PEPTIDE SYNTHETASE MBTE (PEPTIDE SYNTHASE)		Mouse 27, in vitro 29	Macrophage 31	0.007
Rv2536		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	TM	Mouse 27		0.017
Rv2938	drrC	PROBABLE DAUNORUBICIN-DIM-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER DRRC	TM	Mouse 25		0.017
Rv3343c	PPE54	PPE FAMILY PROTEIN		in vitro 28, in vitro 29	Human 38, Human 39	0.006
Rv3478	PPE60, mtb39c	PE FAMILY PROTEIN				0.017
Rv3526		POSSIBLE OXIDOREDUCTASE			Macrophage 31	0.005
Rv3554	fdxB	POSSIBLE ELECTRON TRANSFER PROTEIN FDXB	TM			0.001
Rv3596c	clpC1, clpC	PROBABLE ATP-DEPENDENT PROTEASE ATP-BINDING SUBUNIT CLPC1		Macrophage 23, Macrophage 24, in vitro 28, in vitro 29		0.006
Rv3654c		CONSERVED HYPOTHETICAL PROTEIN			Human 38	0.000
Rv3701c		CONSERVED HYPOTHETICAL PROTEIN		Macrophage 23, Mouse 25		0.039

Rv3707c		CONSERVED HYPOTHETICAL PROTEIN				0.039
Rv3823c	mmpL8	PROBABLE CONSERVED INTEGRAL MEMBRANE TRANSPORT PROTEIN MMPL8	TM	Mouse 27, in vitro 29		0.027

Table 2.10. EXIT identified proteins exported significantly more *in vivo* than *in vitro*. 38 proteins were identified with export upregulated during infection. Proteins are identified by their H37Rv genome designation number, name, and function as annotated in the H37Rv RefSeq genome annotation released January 9 2012.P values were corrected due to multiple testing using the B-H correction to generate q values. Q values less than 0.05, representing less than 5% false discovery rate (FDR), were considered significant. Column 4 identifies all *in silico* predicted export signals: signal peptide (SP, (Petersen *et al.*, 2011)), twin-arginine translocation signal peptide (Tat SP, (McDonough *et al.*, 2008)), lipoprotein signal peptide (Lipo, (Sutcliffe & Harrington, 2004)), or transmembrane domain (TM, (Krogh *et al.*, 2001)). Column 5 identifies all large scale virulence screens that identified a given gene as essential, either *in vitro* or in models of tuberculosis infection. Column 6 identifies all studies where a given gene was identified as transcriptionally induced during infection. Numbers given relate to the references and conditions as described below.

Genes identified as essential for virulence in models of tuberculosis: Macaque: 22 (Dutta *et al.*, 2010), Macrophage: 23 (Rengarajan *et al.*, 2005), 24 (Stewart *et al.*, 2005), Mouse: 25 (Sassetti & Rubin, 2003), 26 (Lamichhane *et al.*, 2005), 27 (Zhang *et al.*, 2013). Genes identified as essential for *in vitro* growth: 28 (Sassetti *et al.*, 2003), 29 (Sassetti *et al.*, 2003). Genes identified as transcriptionally induced during infection: Macrophage: 30 (Dubnau *et al.*, 2002), 31 (Schnappinger *et al.*, 2003), 32 (Rohde *et al.*, 2007b), 33 (BCG) (Rohde *et al.*, 2007b), Mouse: 34 (Dubnau *et al.*, 2005), 35 (Talaat *et al.*, 2007), 36 (60 dpi) (Talaat *et al.*, 2007), 37 (45 dpi) (Talaat *et al.*, 2007), Human: 38 (granuloma) (Rachman *et al.*, 2006a), 39 (distant lung) (Rachman *et al.*, 2006a).

REFERENCES

Genome analysis: comparison of the transport capabilities of several bacteria. http://www.biology.ucsd.edu/~ipaulsen/transport/.

(2013) Correction: A beta-Lactamase based reporter system for ESX dependent protein translocation in mycobacteria. *PLoS One* **8**.

Agbor, T.A. & B.A. McCormick, (2011) Salmonella effectors: important players modulating host cell function during infection. *Cell Microbiol* **13**: 1858-1869.

Angelichio, M.J. & A. Camilli, (2002) In vivo expression technology. *Infect Immun* **70**: 6518-6523.

Bange, F.C., F.M. Collins & W.R.J. Jacobs, (1999) Survival of mice infected with *Mycobacterium smegmatis* containing large DNA fragments from *Mycobacterium tuberculosis*. *Tubercle and Lung Disease* **79**: 171-180.

Banu, S., N. Honore, B. Saint-Joanis, D. Philpott, M.C. Prevost & S.T. Cole, (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol. Micro* **44**: 9-19.

Barinaga, M., (1993) New technique offers a window on bacteria's secret weapons. *Science* **259**: 595.

Becker, D., M. Selbach, C. Rollenhagen, M. Ballmaier, T.F. Meyer, M. Mann & D. Bumann, (2006) Robust Salmonella metabolism limits possibilities for new antimicrobials. *Nature* **440**: 303-307.

Becker, K.W. & E.P. Skaar, (2014) Metal limitation and toxicity at the interface between host and pathogen. *FEMS Microbiol Rev* **38**: 1235-1249.

Bell, C., G.T. Smith, M.J. Sweredoski & S. Hess, (2012) Characterization of the *Mycobacterium tuberculosis* proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. *J Proteome Res* **11**: 119-130.

Bendtsen, J.D., H. Nielsen, D. Widdick, T. Palmer & S. Brunak, (2005) Prediction of twin-arginine signal peptides. *BMC Bioinformatics* **6**: 167.

Boesen, H., B.N. Jensen, T. Wilcke & P. Andersen, (1995) Human T-cell responses to secreted antigen fractions of *Mycobacterium tuberculosis*. *Infect Immun* **63**: 1491-1497.

Bottai, D., T.P. Stinear, P. Supply & R. Brosch, (2014) *Mycobacterial Pathogenomics and Evolution,* p. 824. ASM Press.

Braunstein, M., S.S. Bardarov & W.R.J. Jacobs, (2002) Genetic methods for deciphering virulence determinants of *Mycobacterium tuberculosis*. In: Methods in Enzymology. V.L. Clark & P.M. Bavoil (eds). London: Academic Press, pp. 67-99.

Braunstein, M., B. Espinosa, J. Chan, J.T. Belisle & W.R.J. Jacobs, (2003) SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. *Molecular Microbiology* **48**: 453-464.

Braunstein, M., T.I. Griffin, J.I. Kriakov, S.T. Friedman, N.D. Grindley & W.R. Jacobs, Jr., (2000) Identification of genes encoding exported *Mycobacterium tuberculosis* proteins using a Tn552'phoA in vitro transposition system. *J Bacteriol* **182**: 2732-2740.

Brennan, M.J., G. Delogu, Y. Chen, S. Bardarov, J. Kriakov, M. Alavi & W.R. Jacobs, Jr., (2001) Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun* **69**: 7326-7333.

Brodin, P., Y. Poquet, F. Levillain, I. Peguillet, G. Larrouy-Maumus, M. Gilleron, F. Ewann, T. Christophe, D. Fenistein, J. Jang, M.S. Jang, S.J. Park, J. Rauzier, J.P. Carralot, R. Shrimpton, A. Genovesio, J.A. Gonzalo-Asensio, G. Puzo, C. Martin, R. Brosch, G.R. Stewart, B. Gicquel & O. Neyrolles, (2010) High content phenotypic cell-based visual screen identifies *Mycobacterium tuberculosis* acyltrehalose-containing glycolipids involved in phagosome remodeling. *PLoS Pathog* **6**: e1001100.

Broms, J.E., L. Meyer, K. Sun, M. Lavander & A. Sjostedt, (2012) Unique substrates secreted by the type VI secretion system of *Francisella tularensis* during intramacrophage infection. *PLoS One* **7**: e50473.

Broome-Smith, J.K. & B.G. Spratt, (1986) A vector for the construction of translational fusions to TEM beta-lactamase and the analysis of protein export signals and membrane protein topology. *Gene* **49**: 341-349.

Camacho, L.R., P. Constant, C. Raynaud, M.A. Laneelle, J.A. Triccas, B. Gicquel, M. Daffe & C. Guilhot, (2001) Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. *J Biol Chem* **276**: 19845-19854.

Camacho, L.R., D. Ensergueix, E. Perez, B. Gicquel & C. Guilhot, (1999) Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis [In Process Citation]. *Mol Microbiol* **34**: 257-267.

Cantrell, S.A., M.D. Leavell, O. Marjanovic, A.T. Iavarone, J.A. Leary & L.W. Riley, (2013) Free mycolic acid accumulation in the cell wall of the mce1 operon mutant strain of *Mycobacterium tuberculosis*. *Journal of microbiology* **51**: 619-626.

Casali, N. & L.W. Riley, (2007) A phylogenomic analysis of the Actinomycetales mce operons. *BMC Genomics* **8**: 60.

Chao, J.D., K.G. Papavinasasundaram, X. Zheng, A. Chavez-Steenbock, X. Wang, G.Q. Lee & Y. Av-Gay, (2010) Convergence of Ser/Thr and two-component signaling to coordinate expression of the dormancy regulon in *Mycobacterium tuberculosis*. *J Biol Chem* **285**: 29239-29246.

Chuang, Y.M., N. Bandyopadhyay, D. Rifat, H. Rubin, J.S. Bader & P.C. Karakousis, (2015) Deficiency of the novel exopolyphosphatase Rv1026/PPX2 leads to metabolic downshift and altered cell wall permeability in *Mycobacterium tuberculosis*. *mBio* **6**: e02428.

Claros, M.G. & G. von Heijne, (1994) TopPred II: an improved software for membrane protein structure predictions. *Computer applications in the biosciences : CABIOS* **10**: 685-686.

Converse, S.E., J.D. Mougous, M.D. Leavell, J.A. Leary, C.R. Bertozzi & J.S. Cox, (2003) MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. *Proc Natl Acad Sci U S A* **100**: 6121-6126.

Daleke, M.H., R. Ummels, P. Bawono, J. Heringa, C.M. Vandenbroucke-Grauls, J. Luirink & W. Bitter, (2012) General secretion signal for the mycobacterial type VII secretion pathway. *Proc Natl Acad Sci U S A* **109**: 11342-11347.

Danelishvili, L., L. Babrak, S.J. Rose, J. Everman & L.E. Bermudez, (2014) *Mycobacterium tuberculosis* alters the metalloprotease activity of the COP9 signalosome. *mBio* **5**.

Danelishvili, L., Y. Yamazaki, J. Selker & L.E. Bermudez, (2010) Secreted *Mycobacterium tuberculosis* Rv3654c and Rv3655c proteins participate in the suppression of macrophage apoptosis. *PLoS One* **5**: e10474.

de Souza, G.A., N.A. Leversen, H. Malen & H.G. Wiker, (2011) Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. *Journal of proteomics* **75**: 502-510.

de Souza, G.A. & H.G. Wiker, (2011) A proteomic view of mycobacteria. *Proteomics* **11**: 3118-3127.

Dewoody, R.S., P.M. Merritt & M.M. Marketon, (2013) Regulation of the Yersinia type III secretion system: traffic control. *Frontiers in cellular and infection microbiology* **3**: 4.

Domenech, P., M.B. Reed & C.E. Barry, 3rd, (2005) Contribution of the *Mycobacterium tuberculosis* MmpL protein family to virulence and drug resistance. *Infect Immun* **73**: 3492-3501.

Domenech, P., M.B. Reed, C.S. Dowd, C. Manca, G. Kaplan & C.E. Barry, 3rd, (2004) The role of MmpL8 in sulfatide biogenesis and virulence of *Mycobacterium tuberculosis*. *J Biol Chem* **279**: 21257-21265.

Dubnau, E., J. Chan, V.P. Mohan & I. Smith, (2005) responses of *Mycobacterium tuberculosis* to growth in the mouse lung. *Infect Immun* **73**: 3754-3757.

Dubnau, E., P. Fontan, R. Manganelli, S. Soares-Appel & I. Smith, (2002) *Mycobacterium tuberculosis* genes induced during infection of human macrophages. *Infect Immun* **70**: 2787-2795.

Dubnau, E. & I. Smith, (2003) *Mycobacterium tuberculosis* gene expression in macrophages. *Microbes Infect* **5**: 629-637.

Dutta, N.K., S. Mehra, P.J. Didier, C.J. Roy, L.A. Doyle, X. Alvarez, M. Ratterree, N.A. Be, G. Lamichhane, S.K. Jain, M.R. Lacey, A.A. Lackner & D. Kaushal, (2010) Genetic requirements for the survival of tubercle bacilli in primates. *J Infect Dis* **201**: 1743-1752.

Ehsani, S., C.D. Rodrigues & J. Enninga, (2009) Turning on the spotlight--using light to monitor and characterize bacterial effector secretion and translocation. *Curr Opin Microbiol* **12**: 24-30.

Feltcher, M.E., H.S. Gibbons, L.S. Ligon & M. Braunstein, (2013) Protein export by the mycobacterial SecA2 system is determined by the preprotein mature domain. *J Bacteriol* **195**: 672-681.

Feltcher, M.E., H.P. Gunawardena, K.E. Zulauf, S. Malik, J.E. Griffin, C.M. Sassetti, X. Chen & M. Braunstein, (2015) Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall. *Molecular & cellular proteomics : MCP*.

Feltcher, M.E., J.T. Sullivan & M. Braunstein, (2010) Protein export systems of *Mycobacterium tuberculosis*: novel targets for drug development? *Future Microbiology* **5**: 1581-1597.

Festa, R.A., M.B. Jones, S. Butler-Wu, D. Sinsimer, R. Gerads, W.R. Bishai, S.N. Peterson & K.H. Darwin, (2011) A novel copper-responsive regulon in *Mycobacterium tuberculosis*. *Mol Microbiol* **79**: 133-148.

Flores, A.R., L.M. Parsons & M.S. Pavelka, Jr., (2005) Genetic analysis of the beta-lactamases of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* and susceptibility to beta-lactam antibiotics. *Microbiology* **151**: 521-532.

Forrellad, M.A., M.V. Bianco, F.C. Blanco, J. Nunez, L.I. Klepp, C.L. Vazquez, L. Santangelo Mde, R.V. Rocha, M. Soria, P. Golby, M.G. Gutierrez & F. Bigi, (2013a) Study of the in vivo role of Mce2R, the transcriptional regulator of mce2 operon in *Mycobacterium tuberculosis*. *BMC Microbiol* **13**: 200.

Forrellad, M.A., L.I. Klepp, A. Gioffre, J. Sabio y Garcia, H.R. Morbidoni, M. de la Paz Santangelo, A.A. Cataldi & F. Bigi, (2013b) Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* **4**: 3-66.

Forrellad, M.A., M. McNeil, L. Santangelo Mde, F.C. Blanco, E. Garcia, L.I. Klepp, J. Huff, M. Niederweis, M. Jackson & F. Bigi, (2014) Role of the Mce1 transporter in the lipid homeostasis of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **94**: 170-177.

Fraley, C., A.E. Raftery, L. Scrucca, T.B. Murphy & M. Fop, Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation (mclust). In., pp.

Ge, J., H. Xu, T. Li, Y. Zhou, Z. Zhang, S. Li, L. Liu & F. Shao, (2009) A Legionella type IV effector activates the NF-kappaB pathway by phosphorylating the IkappaB family of inhibitors. *Proc Natl Acad Sci U S A* **106**: 13725-13730.

Gibbons, H.S., F. Wolschendorf, M. Abshire, M. Niederweis & M. Braunstein, (2007) Identification of two *Mycobacterium smegmatis* lipoproteins exported by a SecA2-dependent pathway. *J Bacteriol* **189**: 5090-5100.

Giffin, M.M., R.W. Raab, M. Morganstern & C.D. Sohaskey, (2012) Mutational analysis of the respiratory nitrate transporter NarK2 of *Mycobacterium tuberculosis*. *PLoS One* **7**: e45459.

Gioffre, A., E. Infante, D. Aguilar, M.P. Santangelo, L. Klepp, A. Amadio, V. Meikle, I. Etchechoury, M.I. Romano, A. Cataldi, R.P. Hernandez & F. Bigi, (2005) Mutation in mce operons attenuates *Mycobacterium tuberculosis* virulence. *Microbes Infect* **7**: 325-334.

Glickman, M.S., J.S. Cox & W.R. Jacobs, Jr., (2000) A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. *Mol Cell* **5**: 717-727.

Glover, R.T., J. Kriakov, S.J. Garforth, A.D. Baughn & W.R. Jacobs, Jr., (2007) The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in *Mycobacterium smegmatis*. *J Bacteriol* **189**: 5495-5503.

Goldberg, M., N.K. Saini & S.A. Porcelli, (2014) *Evasion of Innate and Adaptive Immunity by Mycobacterium tuberculosis,* p. 824. ASM Press.

Gomez-Velasco, A., H. Bach, A.K. Rana, L.R. Cox, A. Bhatt, G.S. Besra & Y. Av-Gay, (2013) Disruption of the serine/threonine protein kinase H affects phthiocerol dimycocerosates synthesis in *Mycobacterium tuberculosis*. *Microbiology* **159**: 726-736.

Gomez, M., S. Johnson & M.L. Gennaro, (2000) Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. *Infect Immun* **68**: 2323-2327.

Goude, R., A.G. Amin, D. Chatterjee & T. Parish, (2008) The critical role of embC in *Mycobacterium tuberculosis*. *J Bacteriol* **190**: 4335-4341.

Gouzy, A., Y. Poquet & O. Neyrolles, (2014) Nitrogen metabolism in *Mycobacterium tuberculosis* physiology and virulence. *Nat Rev Microbiol* **12**: 729-737.

Griffin, J.E., J.D. Gawronski, M.A. Dejesus, T.R. Ioerger, B.J. Akerley & C.M. Sassetti, (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**: e1002251.

Gu, S., J. Chen, K.M. Dobos, E.M. Bradbury, J.T. Belisle & X. Chen, (2003) Comprehensive Proteomic Profiling of the Membrane Constituents of a *Mycobacterium tuberculosis* Strain. *Molecular & cellular proteomics : MCP* **2**: 1284-1296.

Gunawardena, H.P., M.E. Feltcher, J.A. Wrobel, S. Gu, M. Braunstein & X. Chen, (2013) Comparison of the membrane proteome of virulent *Mycobacterium tuberculosis* and the attenuated *Mycobacterium bovis* BCG vaccine strain by label-free quantitative proteomics. *J Proteome Res* **12**: 5463-5474.

H. Pages, P.A., R. Gentleman, and S. DebRoy Biostrings: String objects representing biological sequences, and matching algorithms. In., pp.

Hiller, K., A. Grote, M. Scheer, R. Munch & D. Jahn, (2004) PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res* **32**: W375-379.

Hofmann, K. & W. Stoffel, (1993) TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374,**: 166-170.

Houben, E.N., J. Bestebroer, R. Ummels, L. Wilson, S.R. Piersma, C.R. Jimenez, T.H. Ottenhoff, J. Luirink & W. Bitter, (2012) Composition of the type VII secretion system membrane complex. *Mol Microbiol* **86**: 472-484.

Houben, E.N., K.V. Korotkov & W. Bitter, (2014) Take five - Type VII secretion systems of Mycobacteria. *Biochim Biophys Acta* **1843**: 1707-1716.

Hu, Y., R. van der Geize, G.S. Besra, S.S. Gurcha, A. Liu, M. Rohde, M. Singh & A. Coates, (2010) 3-Ketosteroid 9alpha-hydroxylase is an essential factor in the pathogenesis of *Mycobacterium tuberculosis*. *Mol Microbiol* **75**: 107-121.

Isaac, D.T. & R. Isberg, (2014) Master manipulators: an update on *Legionella pneumophila* Icm/Dot translocated substrates and their host targets. *Future Microbiol* **9**: 343-359.

Ivanyi, J., (2014) Function and Potentials of *M. tuberculosis* Epitopes. *Frontiers in immunology* **5**: 107.

Jackson, M., M.R. McNeil & P.J. Brennan, (2013) Progress in targeting cell envelope biogenesis in *Mycobacterium tuberculosis*. *Future Microbiol* **8**: 855-875.

Jain, M., C.J. Petzold, M.W. Schelle, M.D. Leavell, J.D. Mougous, C.R. Bertozzi, J.A. Leary & J.S. Cox, (2007) Lipidomics reveals control of *Mycobacterium tuberculosis* virulence lipids via metabolic coupling. *Proc Natl Acad Sci U S A* **104**: 5133-5138.

Jones, C.M., R.M. Wells, A.V. Madduri, M.B. Renfrow, C. Ratledge, D.B. Moody & M. Niederweis, (2014) Self-poisoning of *Mycobacterium tuberculosis* by interrupting siderophore recycling. *Proc Natl Acad Sci U S A* **111**: 1945-1950.

Jones, D.T., W.R. Taylor & J.M. Thornton, (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **33**: 3038-3049.

Juncker, A.S., H. Willenbrock, G. Von Heijne, S. Brunak, H. Nielsen & A. Krogh, (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein science : a publication of the Protein Society* **12**: 1652-1662.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes & A. Drummond, (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.

Krogh, A., B. Larsson, G. von Heijne & E.L. Sonnhammer, (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.

Kruh, N.A., J. Troudt, A. Izzo, J. Prenni & K.M. Dobos, (2010) Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome in vivo. *PLoS One* **5**: e13938.

Kurtz, S. & M. Braunstein, (2005) Protein secretion and export in *Mycobacterium tuberculosis*. In: *Mycobacterium* molecular biology. T. Parish (ed). Norfolk, UK: Horizon bioscience, pp. 71-138.

La Rosa, V., G. Poce, J.O. Canseco, S. Buroni, M.R. Pasca, M. Biava, R.M. Raju, G.C. Porretta, S. Alfonso, C. Battilocchio, B. Javid, F. Sorrentino, T.R. Ioerger, J.C. Sacchettini, F. Manetti, M. Botta, A. De Logu, E.J. Rubin & E. De Rossi, (2012) MmpL3 is the cellular target of the antitubercular pyrrole derivative BM212. *Antimicrob Agents Chemother* **56**: 324-331.

Laal, S., K.M. Samanich, M.G. Sonnenberg, S. Zolla-Pazner, J.M. Phadtare & J.T. Belisle, (1997) Human humoral responses to antigens of *Mycobacterium tuberculosis*: immunodominance of high-molecular-mass antigens. *Clin Diagn Lab Immunol* **4**: 49-56.

Lamarche, M.G., B.L. Wanner, S. Crepin & J. Harel, (2008) The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. *FEMS Microbiol Rev* **32**: 461-473.

Lamichhane, G., S. Tyagi & W.R. Bishai, (2005) Designer arrays for defined mutant analysis to detect genes essential for survival of *Mycobacterium tuberculosis* in mouse lungs. *Infect Immun* **73**: 2533-2540.

Leversen, N.A., G.A. de Souza, H. Malen, S. Prasad, I. Jonassen & H.G. Wiker, (2009) Evaluation of signal peptide prediction algorithms for identification of mycobacterial signal peptides using sequence data from proteomic methods. *Microbiology* **155**: 2375-2383.

Li, W., A. Upadhyay, F.L. Fontes, E.J. North, Y. Wang, D.C. Crans, A.E. Grzegorzewicz, V. Jones, S.G. Franzblau, R.E. Lee, D.C. Crick & M. Jackson, (2014) Novel insights into the mechanism of inhibition of MmpL3, a target of multiple pharmacophores in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **58**: 6413-6423.

Ligon, L.S., J.D. Hayden & M. Braunstein, (2012) The ins and outs of *Mycobacterium tuberculosis* protein export. *Tuberculosis (Edinb)* **92**: 121-132.

Liu, X., B. Gao, V. Novik & J.E. Galan, (2012) Quantitative Proteomics of Intracellular *Campylobacter jejuni* Reveals Metabolic Reprogramming. *PLoS Pathog* **8**: e1002562.

Long, J.E., M. DeJesus, D. Ward, R.E. Baker, T. Ioerger & C.M. Sassetti, (2015) Identifying essential genes in *Mycobacterium tuberculosis* by global phenotypic profiling. *Methods Mol Biol* **1279**: 79-95.

Lyashchenko, K., R. Colangeli, M. Houde, H. Al Jahdali, D. Menzies & M.L. Gennaro, (1998) Heterogeneous antibody responses in tuberculosis. *Infect Immun* **66**: 3936-3940.

Mahan, M.J., D.M. Heithoff, R.L. Sinsheimer & D.A. Low, (2000) Assessment of bacterial pathogenesis by analysis of gene expression in the host. *Annual review of genetics* **34**: 139-164.

Mahan, M.J., J.M. Slauch & J.J. Mekalanos, (1993) Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* **259**: 686-688.

Mahdavi, A., J. Szychowski, J.T. Ngo, M.J. Sweredoski, R.L. Graham, S. Hess, O. Schneewind, S.K. Mazmanian & D.A. Tirrell, (2014) Identification of secreted bacterial proteins by noncanonical amino acid tagging. *Proc Natl Acad Sci U S A* **111**: 433-438.

Malen, H., F.S. Berven, K.E. Fladmark & H.G. Wiker, (2007) Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics* **7**: 1702-1718.

Malen, H., G.A. De Souza, S. Pathak, T. Softeland & H.G. Wiker, (2011) Comparison of membrane proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra strains. *BMC Microbiol* **11**: 18.

Malik-Kale, P., C.E. Jolly, S. Lathrop, S. Winfree, C. Luterbach & O. Steele-Mortimer, (2011) Salmonella - at home in the host cell. *Frontiers in microbiology* **2**: 125.

Marjanovic, O., T. Miyata, A. Goodridge, L.V. Kendall & L.W. Riley, (2010) Mce2 operon mutant strain of *Mycobacterium tuberculosis* is attenuated in C57BL/6 mice. *Tuberculosis (Edinb)* **90**: 50-56.

Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S.T. Cole & R. Brosch, (2004) Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* **150**: 483-496.

Mawuenyega, K.G., C.V. Forst, K.M. Dobos, J.T. Belisle, J. Chen, E.M. Bradbury, A.R. Bradbury & X. Chen, (2005) *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. *Mol Biol Cell* **16**: 396-404.

McCann, J.R., J.A. McDonough, M.S. Pavelka & M. Braunstein, (2007) Beta-lactamase can function as a reporter of bacterial protein export during *Mycobacterium tuberculosis* infection of host cells. *Microbiology* **153**: 3350-3359.

McCann, J.R., J.A. McDonough, J.T. Sullivan, M.E. Feltcher & M. Braunstein, (2011) Genome-wide identification of *Mycobacterium tuberculosis* exported proteins with roles in intracellular growth. *J Bacteriol* **193**: 854-861.

McDonough, J.A., J.R. McCann, E.M. Tekippe, J.S. Silverman, N.W. Rigel & M. Braunstein, (2008) Identification of functional Tat signal sequences in *Mycobacterium tuberculosis* proteins. *J Bacteriol* **190**: 6428-6438.

Mohn, W.W., R. van der Geize, G.R. Stewart, S. Okamoto, J. Liu, L. Dijkhuizen & L.D. Eltis, (2008) The actinobacterial mce4 locus encodes a steroid transporter. *J Biol Chem* **283**: 35368-35374.

Nakai, K. & P. Horton, (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**: 34-36.

Neyrolles, O., F. Wolschendorf, A. Mitra & M. Niederweis, (2015) Mycobacteria, metals, and the macrophage. *Immunological reviews* **264**: 249-263.

Niederweis, M., (2008) Nutrient acquisition by mycobacteria. *Microbiology* **154**: 679-692.

Ohol, Y.M., D.H. Goetz, K. Chan, M.U. Shiloh, C.S. Craik & J.S. Cox, (2010) *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host Microbe* **7**: 210-220.

Orme, I. & M. Gonzalez-Juarrero, (2007) Animal models of *M. tuberculosis* Infection. *Current protocols in microbiology* **Chapter 10**: Unit 10A 15.

Pandey, A.K. & C.M. Sassetti, (2008) Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A* **105**: 4376-4380.

Papavinasasundaram, K.G., B. Chan, J.H. Chung, M.J. Colston, E.O. Davis & Y. Av-Gay, (2005) Deletion of the *Mycobacterium tuberculosis* pknH gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice. *J Bacteriol* **187**: 5751-5760.

Pavelka, M.S., Jr. & W.R. Jacobs, Jr., (1999) Comparison of the construction of unmarked deletion mutations in *Mycobacterium smegmatis*, *Mycobacterium bovis* bacillus Calmette-Guerin, and *Mycobacterium tuberculosis* H37Rv by allelic exchange. *J Bacteriol* **181**: 4780-4789.

Pechous, R.D., V. Sivaraman, P.A. Price, N.M. Stasulli & W.E. Goldman, (2013) Early host cell targets of *Yersinia pestis* during primary pneumonic plague. *PLoS Pathog* **9**: e1003679.

Petersen, T.N., S. Brunak, G. von Heijne & H. Nielsen, (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**: 785-786.

Pieper, R., Q. Zhang, P.P. Parmar, S.T. Huang, D.J. Clark, H. Alami, A. Donohue-Rolfe, R.D. Fleischmann, S.N. Peterson & S. Tzipori, (2009) The *Shigella dysenteriae* serotype 1 proteome, profiled in the host intestinal environment, reveals major metabolic modifications and increased expression of invasive proteins. *Proteomics* **9**: 5029-5045.

Poce, G., R.H. Bates, S. Alfonso, M. Cocozza, G.C. Porretta, L. Ballell, J. Rullas, F. Ortega, A. De Logu, E. Agus, V. La Rosa, M.R. Pasca, E. De Rossi, B. Wae, S.G. Franzblau, F. Manetti, M. Botta & M. Biava, (2013) Improved BM212 MmpL3 inhibitor analogue shows efficacy in acute murine model of tuberculosis infection. *PLoS One* **8**: e56980.

Punta, M., L.R. Forrest, H. Bigelow, A. Kernytsky, J. Liu & B. Rost, (2007) Membrane protein prediction methods. *Methods* **41**: 460-474.

Rachman, H., M. Strong, U. Schaible, J. Schuchhardt, K. Hagens, H. Mollenkopf, D. Eisenberg & S.H. Kaufmann, (2006a) *Mycobacterium tuberculosis* gene expression profiling within the context of protein networks. *Microbes Infect* **8**: 747-757.

Rachman, H., M. Strong, T. Ulrichs, L. Grode, J. Schuchhardt, H. Mollenkopf, G.A. Kosmiadi, D. Eisenberg & S.H. Kaufmann, (2006b) Unique transcriptome signature of *Mycobacterium tuberculosis* in pulmonary tuberculosis. *Infect Immun* **74**: 1233-1242.

Rayasam, G.V., (2014) MmpL3 a potential new target for development of novel anti-tuberculosis drugs. *Expert Opin Ther Targets* **18**: 247-256.

Reddy, P.V., R.V. Puri, P. Chauhan, R. Kar, A. Rohilla, A. Khera & A.K. Tyagi, (2013) Disruption of mycobactin biosynthesis leads to attenuation of *Mycobacterium tuberculosis* for growth and virulence. *J Infect Dis* **208**: 1255-1265.

Remuinan, M.J., E. Perez-Herran, J. Rullas, C. Alemparte, M. Martinez-Hoyos, D.J. Dow, J. Afari, N. Mehta, J. Esquivias, E. Jimenez, F. Ortega-Muro, M.T. Fraile-Gabaldon, V.L. Spivey, N.J. Loman, M.J. Pallen, C. Constantinidou, D.J. Minick, M. Cacho, M.J. Rebollo-Lopez, C. Gonzalez, V. Sousa, I. Angulo-Barturen, A. Mendoza-Losana, D. Barros, G.S. Besra, L. Ballell & N. Cammack, (2013) Tetrahydropyrazolo[1,5-a]pyrimidine-3-carboxamide and N-benzyl-6',7'-dihydrospiro[piperidine-4,4'-thieno[3,2-c]pyran] analogues with bactericidal efficacy against *Mycobacterium tuberculosis* targeting MmpL3. *PLoS One* **8**: e60933.

Rengarajan, J., B.R. Bloom & E.J. Rubin, (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* **102**: 8327-8332.

Rifat, D., D.A. Belchis & P.C. Karakousis, (2014) senX3-independent contribution of regX3 to *Mycobacterium tuberculosis* virulence. *BMC Microbiol* **14**: 265.

Rifat, D., W.R. Bishai & P.C. Karakousis, (2009) Phosphate depletion: a novel trigger for *Mycobacterium tuberculosis* persistence. *J Infect Dis* **200**: 1126-1135.

Rifat, D. & P.C. Karakousis, (2014) Differential regulation of the two-component regulatory system senX3-regX3 in *Mycobacterium tuberculosis*. *Microbiology* **160**: 1125-1133.

Robbins, N., S.E. Koch, M. Tranter & J. Rubinstein, (2012) The history and future of probenecid. *Cardiovascular toxicology* **12**: 1-9.

Rodriguez, J.E., A.S. Ramirez, L.P. Salas, C. Helguera-Repetto, J. Gonzalez-y-Merchand, C.Y. Soto & R. Hernandez-Pando, (2013) Transcription of genes involved in sulfolipid and polyacyltrehalose biosynthesis of *Mycobacterium tuberculosis* in experimental latent tuberculosis infection. *PLoS One* **8**: e58378.

Rohde, K., R.M. Yates, G.E. Purdy & D.G. Russell, (2007a) *Mycobacterium tuberculosis* and the environment within the phagosome. *Immunological reviews* **219**: 37-54.

Rohde, K.H., R.B. Abramovitch & D.G. Russell, (2007b) *Mycobacterium tuberculosis* invasion of macrophages: linking bacterial gene expression to environmental cues. *Cell Host Microbe* **2**: 352-364.

Rose, R.W., T. Bruser, J.C. Kissinger & M. Pohlschroder, (2002) Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol Microbiol* **45**: 943-950.

Rosenberger, T., J.K. Brulle & P. Sander, (2012) A beta-Lactamase based reporter system for ESX dependent protein translocation in mycobacteria. *PLoS One* **7**: e35453.

Rosenkrands, I., A. King, K. Weldingh, M. Moniatte, E. Moertz & P. Andersen, (2000a) Towards the proteome of *Mycobacterium tuberculosis*. *Electrophoresis* **21**: 3740-3756.

Rosenkrands, I., K. Weldingh, S. Jacobsen, C.V. Hansen, W. Florio, I. Gianetri & P. Andersen, (2000b) Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. *Electrophoresis* **21**: 935-948.

Rowland, J.L. & M. Niederweis, (2013) A multicopper oxidase is required for copper resistance in *Mycobacterium tuberculosis*. *J Bacteriol* **195**: 3724-3733.

Russell, D.G., C.E. Barry, 3rd & J.L. Flynn, (2010) Tuberculosis: what we don't know can, and does, hurt us. *Science* **328**: 852-856.

Saleh, M.T., M. Fillon, P.J. Brennan & J.T. Belisle, (2001) Identification of putative exported/secreted proteins in prokaryotic proteomes. *Gene* **269**: 195-204.

Samanich, K.M., J.T. Belisle, M.G. Sonnenberg, M.A. Keen, S. Zolla-Pazner & S. Laal, (1998) Delineation of human antibody responses to culture filtrate antigens of *Mycobacterium tuberculosis*. *J Infect Dis* **178**: 1534-1538.

Samanich, K.M., M.A. Keen, V.D. Vissa, J.D. Harder, J.S. Spencer, J.T. Belisle, S. Zolla-Pazner & S. Laal, (2000) Serodiagnostic potential of culture filtrate antigens of *Mycobacterium tuberculosis*. *Clin Diagn Lab Immunol* **7**: 662-668.

Sandhu, P. & Y. Akhter, (2015) The internal gene duplication and interrupted coding sequences in the MmpL genes of *Mycobacterium tuberculosis*: Towards understanding the multidrug transport in an evolutionary perspective. *International journal of medical microbiology : IJMM*.

Santangelo Mde, L., F. Blanco, E. Campos, M. Soria, M.V. Bianco, L. Klepp, A. Alito, O. Zabal, A. Cataldi & F. Bigi, (2009) Mce2R from *Mycobacterium tuberculosis* represses the expression of the mce2 operon. *Tuberculosis (Edinb)* **89**: 22-28.

Santangelo, M.P., J. Goldstein, A. Alito, A. Gioffre, K. Caimi, O. Zabal, M. Zumarraga, M.I. Romano, A.A. Cataldi & F. Bigi, (2002) Negative transcriptional regulation of the mce3 operon in *Mycobacterium tuberculosis*. *Microbiology* **148**: 2997-3006.

Sassetti, C.M., D.H. Boyd & E.J. Rubin, (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**: 77-84.

Sassetti, C.M. & E.J. Rubin, (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**: 12989-12994.

Schmidt, F., S.S. Scharf, P. Hildebrandt, M. Burian, J. Bernhardt, V. Dhople, J. Kalinka, M. Gutjahr, E. Hammer & U. Volker, (2010) Time-resolved quantitative proteome profiling of host-pathogen interactions: the response of *Staphylococcus aureus* RN1HG to internalisation by human airway epithelial cells. *Proteomics* **10**: 2801-2811.

Schmidt, F. & U. Volker, (2011) Proteome analysis of host-pathogen interactions: Investigation of pathogen responses to the host cell environment. *Proteomics* **11**: 3203-3211.

Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan & G.K. Schoolnik, (2003) Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *J Exp Med* **198**: 693-704.

Seidel, M., L.J. Alderwick, H. Sahm, G.S. Besra & L. Eggeling, (2007) Topology and mutational analysis of the single Emb arabinofuranosyltransferase of *Corynebacterium glutamicum* as a model of Emb proteins of *Mycobacterium tuberculosis*. *Glycobiology* **17**: 210-219.

Selengut, J.D., D.H. Haft, T. Davidsen, A. Ganapathy, M. Gwinn-Giglio, W.C. Nelson, A.R. Richter & O. White, (2007) TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res* **35**: D260-264.

Senaratne, R.H., B. Sidders, P. Sequeira, G. Saunders, K. Dunphy, O. Marjanovic, J.R. Reader, P. Lima, S. Chan, S. Kendall, J. McFadden & L.W. Riley, (2008) *Mycobacterium tuberculosis* strains disrupted in mce3 and mce4 operons are attenuated in mice. *Journal of medical microbiology* **57**: 164-170.

Sharma, K., M. Gupta, M. Pathak, N. Gupta, A. Koul, S. Sarangi, R. Baweja & Y. Singh, (2006) Transcriptional control of the mycobacterial embCAB operon by PknH through a regulatory protein, EmbR, in vivo. *J Bacteriol* **188**: 2936-2944.

Shi, L., C.D. Sohaskey, B.D. Kana, S. Dawes, R.J. North, V. Mizrahi & M.L. Gennaro, (2005) Changes in energy metabolism of *Mycobacterium tuberculosis* in mouse lung and under in vitro conditions affecting aerobic respiration. *Proc Natl Acad Sci U S A* **102**: 15629-15634.

Shi, X. & K.H. Darwin, (2015) Copper homeostasis in *Mycobacterium tuberculosis*. *Metallomics : integrated biometal science*.

Shi, X., R.A. Festa, T.R. Ioerger, S. Butler-Wu, J.C. Sacchettini, K.H. Darwin & M.I. Samanovic, (2014) The copper-responsive RicR regulon contributes to *Mycobacterium tuberculosis* virulence. *mBio* **5**.

Singh, K.K., X. Zhang, A.S. Patibandla, P. Chien, Jr. & S. Laal, (2001) Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. *Infect Immun* **69**: 4185-4191.

Sohaskey, C.D. & L.G. Wayne, (2003) Role of narK2X and narGHJI in hypoxic upregulation of nitrate reduction by *Mycobacterium tuberculosis*. *J Bacteriol* **185**: 7247-7256.

Sonnhammer, E.L., G. von Heijne & A. Krogh, (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**: 175-182.

Srivastava, V., C. Rouanet, R. Srivastava, B. Ramalingam, C. Locht & B.S. Srivastava, (2007) Macrophage-specific *Mycobacterium tuberculosis* genes: identification by green fluorescent protein and kanamycin resistance selection. *Microbiology* **153**: 659-666.

Stewart, G.R., J. Patel, B.D. Robertson, A. Rae & D.B. Young, (2005) Mycobacterial mutants with defective control of phagosomal acidification. *PLoS Pathog* **1**: 269-278.

Stover, C.K., V.F. de la Cruz, T.R. Fuerst, J.E. Burlein, L.A. Benson, L.T. Bennett, G.P. Bansal, J.F. Young, M.H. Lee & G.F. Hatfull, (1991) New use of BCG for recombinant vaccines. *Nature* **351**: 456-460.

Sutcliffe, I.C. & D.J. Harrington, (2004) Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiol Rev* **28**: 645-659.

Sutcliffe, I.C. & R.R. Russell, (1995) Lipoproteins of gram-positive bacteria. *J Bacteriol* **177**: 1123-1128.

Tahlan, K., R. Wilson, D.B. Kastrinsky, K. Arora, V. Nair, E. Fischer, S.W. Barnes, J.R. Walker, D. Alland, C.E. Barry, 3rd & H.I. Boshoff, (2012) SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **56**: 1797-1809.

Talaat, A.M., R. Lyons, S.T. Howard & S.A. Johnston, (2004) The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. *Proc Natl Acad Sci U S A* **101**: 4602-4607.

Talaat, A.M., S.K. Ward, C.W. Wu, E. Rondon, C. Tavano, J.P. Bannantine, R. Lyons & S.A. Johnston, (2007) Mycobacterial bacilli are metabolically active during chronic tuberculosis in murine lungs: insights from genome-wide transcriptional profiling. *J Bacteriol* **189**: 4265-4274.

Tanrikulu, I.C., E. Schmitt, Y. Mechulam, W.A. Goddard, 3rd & D.A. Tirrell, (2009) Discovery of *Escherichia coli* methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc Natl Acad Sci U S A* **106**: 15285-15290.

Taylor, P.D., C.P. Toseland, T.K. Attwood & D.R. Flower, (2006) TATPred: a Bayesian method for the identification of twin arginine translocation pathway signal sequences. *Bioinformation* **1**: 184-187.

Tekaia, F., S.V. Gordon, T. Garnier, R. Brosch, B.G. Barrell & S.T. Cole, (1999) Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis* **79**: 329-342.

Telenti, A., W.J. Philipp, S. Sreevatsan, C. Bernasconi, K.E. Stockbauer, B. Wieles, J.M. Musser & W.R. Jacobs, Jr., (1997) The *emb* operon, a gene cluster of *Mycobacterium tuberculosis* involved in resistance to ethambutol. *Nat Med* **3**: 567-570.

Timm, J., F.A. Post, L.G. Bekker, G.B. Walther, H.C. Wainwright, R. Manganelli, W.T. Chan, L. Tsenova, B. Gold, I. Smith, G. Kaplan & J.D. McKinney, (2003) Differential expression of iron-, carbon-, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. *Proc Natl Acad Sci U S A* **100**: 14321-14326.

Tischler, A.D., R.L. Leistikow, M.A. Kirksey, M.I. Voskuil & J.D. McKinney, (2013) *Mycobacterium tuberculosis* requires phosphate-responsive gene regulation to resist host immunity. *Infect Immun* **81**: 317-328.

Tomich, M., P.J. Planet & D.H. Figurski, (2007) The tad locus: postcards from the widespread colonization island. *Nat Rev Microbiol* **5**: 363-375.

Twine, S.M., N.C. Mykytczuk, M.D. Petit, H. Shen, A. Sjostedt, J. Wayne Conlan & J.F. Kelly, (2006) In vivo proteomic analysis of the intracellular bacterial pathogen, *Francisella tularensis*, isolated from mouse spleen. *Biochem Biophys Res Commun* **345**: 1621-1633.

van der Woude, A.D., E.J. Stoop, M. Stiess, S. Wang, R. Ummels, G. van Stempvoort, S.R. Piersma, A. Cascioferro, C.R. Jimenez, E.N. Houben, J. Luirink, J. Pieters, A.M. van der Sar & W. Bitter, (2014) Analysis of SecA2-dependent substrates in *Mycobacterium marinum* identifies protein kinase G (PknG) as a virulence effector. *Cell Microbiol* **16**: 280-295.

Varela, C., D. Rittmann, A. Singh, K. Krumbach, K. Bhatt, L. Eggeling, G.S. Besra & A. Bhatt, (2012) MmpL genes are associated with mycolic acid metabolism in mycobacteria and corynebacteria. *Chem Biol* **19**: 498-506.

von Heijne, G., (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol* **225**: 487-494.

von Heijne, G. & Y. Gavel, (1988) Topogenic signals in integral membrane proteins. *Eur J Biochem* **174**: 671-678.

Voskuil, M.I., D. Schnappinger, R. Rutherford, Y. Liu & G.K. Schoolnik, (2004) Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. *Tuberculosis (Edinb)* **84**: 256-262.

Wagner, J.M., T.J. Evans, J. Chen, H. Zhu, E.N. Houben, W. Bitter & K.V. Korotkov, (2013) Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. *Journal of structural biology* **184**: 115-128.

Wang, M., C. Herrmann, M. Simonovic, D. Szklarczyk & C. von Mering, (2015) Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues and cell-lines. *Proteomics*.

Wang, M., M. Weiss, M. Simonovic, G. Haertinger, S.P. Schrimpf, M.O. Hengartner & C. von Mering, (2012) PaxDb, a database of protein abundance averages across all three domains of life. *Molecular & cellular proteomics : MCP* **11**: 492-500.

Ward, S.K., B. Abomoelak, E.A. Hoye, H. Steinberg & A.M. Talaat, (2010) CtpV: a putative copper exporter required for full virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* **77**: 1096-1110.

Weiner, J., 3rd & S.H. Kaufmann, (2014) Recent advances towards tuberculosis control: vaccines and biomarkers. *Journal of internal medicine* **275**: 467-480.

Wells, R.M., C.M. Jones, Z. Xi, A. Speer, O. Danilchanka, K.S. Doornbos, P. Sun, F. Wu, C. Tian & M. Niederweis, (2013) Discovery of a siderophore export system essential for virulence of *Mycobacterium tuberculosis*. *PLoS Pathog* **9**: e1003120.

Wolfe, L.M., S.B. Mahaffey, N.A. Kruh & K.M. Dobos, (2010) Proteomic definition of the cell wall of *Mycobacterium tuberculosis*. *J Proteome Res* **9**: 5816-5826.

Xia, Q., T. Wang, F. Taub, Y. Park, C.A. Capestany, R.J. Lamont & M. Hackett, (2007) Quantitative proteomics of intracellular *Porphyromonas gingivalis*. *Proteomics* **7**: 4323-4337.

Xiong, Y., M.J. Chalmers, F.P. Gao, T.A. Cross & A.G. Marshall, (2005) Identification of *Mycobacterium tuberculosis* H37Rv integral membrane proteins by one-dimensional gel electrophoresis and liquid chromatography electrospray ionization tandem mass spectrometry. *J Proteome Res* **4**: 855-861.

Yang, Y., M. Hu, K. Yu, X. Zeng & X. Liu, (2015) Mass spectrometry-based proteomic approaches to study pathogenic bacteria-host interactions. *Protein & cell*.

Zhang, Y.J., M.C. Reddy, T.R. Ioerger, A.C. Rothchild, V. Dartois, B.M. Schuster, A. Trauner, D. Wallis, S. Galaviz, C. Huttenhower, J.C. Sacchettini, S.M. Behar & E.J. Rubin, (2013) Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. *Cell* **155**: 1296-1308.

Zolotarev, A.S., M. Unnikrishnan, B.E. Shmukler, J.S. Clark, D.H. Vandorpe, N. Grigorieff, E.J. Rubin & S.L. Alper, (2008) Increased sulfate uptake by *E. coli* overexpressing the SLC26-related SulP protein Rv1739c from *Mycobacterium tuberculosis. Comparative biochemistry and physiology. Part A, Molecular & integrative physiology* **149**: 255-266.

CHAPTER 3: AN ORPHANED MCE-ASSOCIATED PROTEIN OF *MYCOBACTERIUM TUBERCULOSIS* IS A VIRULENCE FACTOR THAT STABILIZES MCE TRANSPORTERS[2]

Introduction

Mycobacterium tuberculosis is a human pathogen with a significant impact on world health. Current estimates suggest that 2 billion people worldwide have been infected with *M. tuberculosis* and 1.5 million people die per year from tuberculosis (World Health Organization, 2014). Inside the host, *M. tuberculosis* survives and grows inside macrophages (Rohde *et al.*, 2007). Bacterial exported proteins are proteins that are transported out of cytoplasm to the bacterial cytoplasmic membrane, cell wall, or are fully released into the host environment. Many intracellular pathogens, including *M. tuberculosis*, survive in macrophages with the help of exported proteins (Ligon *et al.*, 2012; Hicks & Galan, 2013; Isaac & Isberg, 2014). Because of their extracytoplasmic location, exported proteins of pathogens are ideally positioned for host interactions and for roles in controlling the immune response, surviving in macrophages, and acquiring nutrients for intracellular survival (Forrellad *et al.*, 2013; McCann, 2009). While exported proteins are known to play a critical role in *M. tuberculosis* virulence, up to 69% of *M. tuberculosis* exported proteins have no assigned function (Appendix III). To better understand how *M. tuberculosis* interacts with the host and causes disease, it is critical to identify the function of unknown exported proteins.

[2] Adapted for this dissertation from: Perkowski EF, McCann JR, Sullivan JT, Malik S, Allen IC, Hayden JD, Godfrey V, Braunstein, M. An orphaned Mce-associated protein of *Mycobacterium tuberculosis* is a virulence factor that stabilizes Mce transporters. Under review.

In a previous study using a transposon carrying a β-lactamase reporter of export, the *M. tuberculosis* Rv0199 protein was identified as an exported protein (McCann *et al.*, 2011). Rv0199 is a 24 kDa protein that is predicted to be an integral membrane protein with an N-terminal transmembrane domain (Krogh *et al.*, 2001). The site of transposon insertion is consistent with the majority of the protein being exported to the periplasmic/cell wall side of the membrane (Figure 3.1A) (McCann *et al.*, 2011). By testing the *rv0199* transposon mutant in cultured macrophages, the Rv0199 protein was shown to be required for intracellular growth of *M. tuberculosis* (McCann *et al.*, 2011). Further, the Rv0199 protein of *Mycobacterium bovis* BCG is predicted to be important for growth in macrophages (Stewart *et al.*, 2005). The *rv0199* gene is a core mycobacterial gene (Marmiesse *et al.*, 2004), which means that it is highly conserved throughout pathogenic and non-pathogenic mycobacterial species but not conserved outside of actinomycetes. However, the function of Rv0199 is not clear. Rv0199 is annotated as a membrane protein of unknown function (Lew *et al.*, 2011), although there is limited sequence homology between the C-terminal region of Rv0199 and the C-terminal region of Mce-associated (Mas) proteins of *M. tuberculosis* (Casali & Riley, 2007). Mas proteins are named for the genomic location of their corresponding genes, which are linked to *mce* operons encoding Mce transporter systems (Casali & Riley, 2007). Besides eight *mas* genes linked to *mce* operons, *M. tuberculosis* has five Open Reading Frames (ORFs) encoding proteins with low levels of homology to Mas proteins that are scattered elsewhere in the genome (Figure 3.1B). Rv0199 is one of these potential orphaned Mas proteins (Casali & Riley, 2007).

Mce transporters are multi-protein complexes considered to be functionally analogous to ABC transporters (Casali & Riley, 2007). The four Mce transporter systems in *M. tuberculosis*

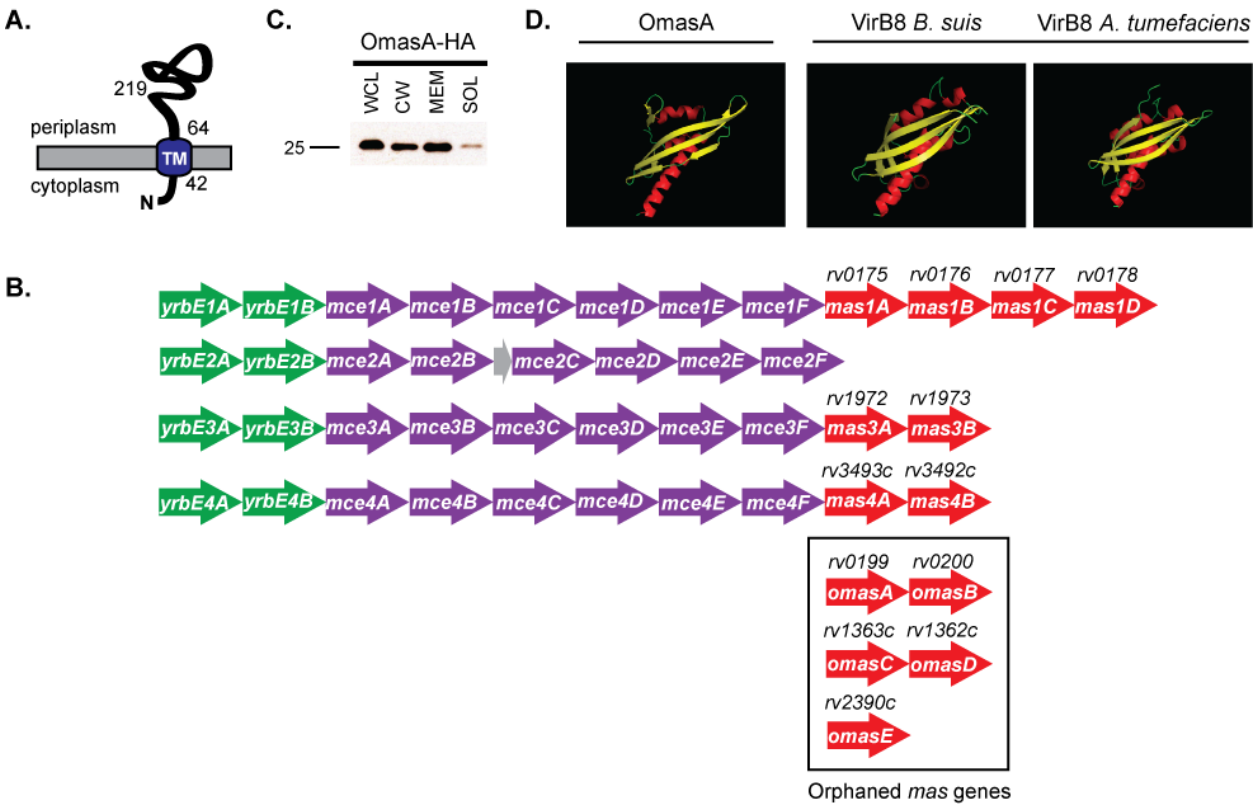

Figure 3.1. Rv0199 (OmasA) is a transmembrane protein predicted to be a Mce-associated protein. **A.** OmasA is predicted to have a single N-terminal transmembrane domain (TM) at amino acid 42-64 (Krogh *et al.*, 2001) and C-terminal domain exposed to the cell wall side of the membrane (McCann *et al.*, 2011). **B.** OmasA is predicted to be a Mce-associated (Mas) protein, however, *rv0199* is not located in a *mce* operon. *Mce* operons are typically organized by two *yrbE* genes upstream (green), six *mce* genes (purple) and most have pairs of *mas* genes (red) downstream. Genes encoding putative *orphaned mas* genes are boxed. Genes encoding Omas proteins are distinguished by being distally located from *mce* operons (Casali & Riley, 2007). The *mce2* operon additionally contains a small predicted pseudogene (grey). **C.** The $omasA_{mtb}$ gene was engineered in frame with an HA tag and expressed in *M. smegmatis*. Cells were lysed to generate whole cell lysates (WCL) and fractionated by differential ultracentrifugation into cell wall (CW), cell membrane (MEM), and cytoplasmic containing soluble (SOL) fractions. Results are representative of at least three independent replicates. **D.** Phyre 2, an online structural prediction program, predicted with high confidence (96%) that OmasA forms a NTF2-like fold, as found in the association domain of the calcium/calmodulin-dependent protein kinase type II alpha subunit (CAMKIIγ). Ribbon diagrams shown represent the Phyre 2 predicted structures of OmasA colored by secondary structure in Pymol. Ribbon diagrams representing the solved crystal structures of VirB8 from *Brucella suis* and *Agrobacterium tumefaciens* are shown for comparison. Alpha helices are colored in red, Beta-strands in yellow, and turns in green.

all play roles in virulence (Gioffre *et al.*, 2005; Shimono *et al.*, 2003; Marjanovic *et al.*, 2010; Lima *et al.*, 2007; McCann *et al.*, 2011; Senaratne *et al.*, 2008; Pandey & Sassetti, 2008) and are thought to function in lipid uptake. The best characterized Mce transporter is Mce4. Mce4 is required for cholesterol uptake (Pandey & Sassetti, 2008; Mohn *et al.*, 2008), cholesterol being an important nutrient during *M. tuberculosis* infection (Pandey & Sassetti, 2008). Emerging evidence suggests that Mce1 is responsible for import of mycolic acids, long chain fatty acids characteristic of mycobacteria (Forrellad *et al.*, 2014; Cantrell *et al.*, 2013). Each Mce system is composed of two YrbE proteins with similarity to ABC transporter permeases and six Mce proteins that are considered functionally similar to substrate binding proteins of ABC transporters (Casali & Riley, 2007). Additionally, Mce transporters are thought to share a common ATPase, MceG. Interestingly, the *mceG* gene is not located near any *mce* operon (Casali & Riley, 2007; Joshi *et al.*, 2006). Nearly all *mce* operons also contain genes encoding Mas proteins (Figure 3.1B). Unlike the YrbE and Mce components, Mas proteins share no analogous features with ABC transporter components. Mas proteins are speculated to have a role in Mce transporter systems, but this idea is solely based on the genomic location of *mas* genes. To date, there have been no functional studies of any Mas protein. Consequently, the function of potential orphaned Mas proteins such as Rv0199, whose genes are distal to *mce* operons, is even less clear.

Here, we further characterized the role of Rv0199 in *M. tuberculosis* virulence using a low dose aerosol model of murine infection. We additionally showed that Rv0199 has a role in Mce lipid transport, leading us to rename Rv0199 as OmasA (orphaned Mce-associated protein A), and we demonstrated a role for OmasA in stabilizing Mce transporter complexes. The stabilization function of OmasA may be analogous to the role of VirB8 in stabilizing Type IV

secretion systems, as structural similarities between Mas proteins and VirB8 proteins are predicted by the Phyre 2 program (Kelley & Sternberg, 2009). Our results provide important functional information about an exported protein with a previously unknown role in virulence and provide the first evidence for any Mas protein functioning with Mce transporters. Finally, our results suggest that OmasA, and possibly other Mas proteins as well, have a structural role important for the stability and/or assembly of Mce transporters.

Results

OmasA is important for murine infection

Previous studies revealed a transposon insertion in the *M. tuberculosis rv0199* gene, hereafter referred to as *omasA,* results in a growth defect in resting murine bone-marrow derived macrophages (McCann *et al.*, 2011). To further explore the role of the OmasA protein in *M. tuberculosis* infection, we evaluated the course of murine infection with the *omasA* transposon mutant (*omasA::tn*) and compared it to infection with an *omasA*^{WT} strain, hereon referred to as wild type (WT) (McCann *et al.*, 2011). Groups of C57BL/6 mice were infected by low dose aerosol with WT, *omasA::tn*, or a complemented *omasA::tn +omasA* strain. Mice infected with the *omasA* mutant had lower bacterial burden in the lungs at 6, 13, and 20 days post-infection compared to WT (Figure 3.2A). However, by 127 days post-infection there was no longer any difference in bacterial burden in the lungs of mice infected with the *omasA* mutant strain as compared to WT infected mice (Figure 3.2A). We also observed reduced bacterial burden in the spleen and liver of the *omasA* mutant infected mice compared to WT at 20 days post-infection and, like the burden in the lungs, the number of *omasA* mutant bacteria reached equivalent levels to WT by 127 days post-infection (data not shown). Importantly, all defects in bacterial burden

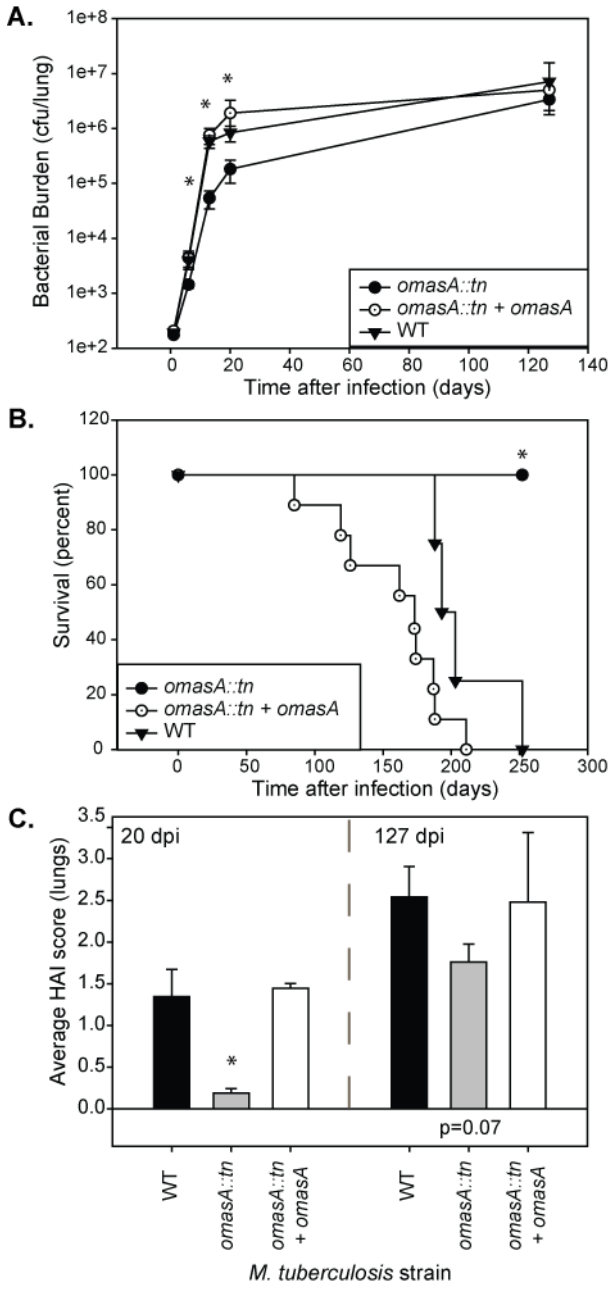

Figure 3.2. OmasA is required for early growth and virulence during murine infection. C57BL/6 mice were infected with a low dose aerosol of WT, *omasA::tn*, or *omasA::tn* +*omasA* complemented strains. **A.** Groups of four mice per strain were sacrificed at various days post infection (dpi) and bacterial burden (colony forming units, cfu) was assessed by plating from organ homogenates. **B.** Groups of mice were monitored for survival. **C.** A single lung lobe from mice sacrificed for bacterial burden was fixed and H&E stained for histology. Average histological activity index (HAI) scores were determined by an experienced blinded reviewer. * indicates $p<0.05$ as compared to WT. tn indicates transposon insertion. Error bars represent standard deviation. Results are representative of two independent experiments comparing WT (MBTB178), *omasA::tn* (MBTB319), and *omasA::tn* + *omasA* (MBTB320).

were fully restored in mice infected with the complemented strain. This data indicates that OmasA is important for early exponential phase growth in the mouse model of infection.

We also assessed long-term survival of mice infected with the strains described above. Because the bacterial burden in mice infected with the *omasA* mutant caught up later in infection, we were somewhat surprised to observe that *omasA* mutant infected animals survived significantly longer (>250 days) compared to WT (193 days average) and the complemented strain (173 days average) (Figure 3.2B). The complemented strain not only alleviated the attenuated phenotype of the *omasA* mutant, but also appeared to potentially accelerate time to death in comparison to WT *M. tuberculosis* (p=0.05). The behavior of the complemented strain may be due to non-physiological levels of OmasA, as the gene is expressed off the constitutive *hsp60* promoter on a multi-copy complementation plasmid. In a separate experiment, when *omasA* mutant infected mice were followed until they succumbed to the infection, the *omasA* mutant infected mice survived almost 50% longer than WT infected mice (data not shown).

H&E stained lung sections demonstrated that mice infected with the *omasA* mutant displayed reduced inflammatory infiltration and increased open alveolar spaces in comparison to WT infected mice. The *omasA* mutant showed this reduced histopathology in both early (Day 20) and late (Day 127) timepoints, and the phenotypes were fully restored in the complemented strain (Figure 3.3). Blinded scoring of these sections demonstrated that the *omasA* mutant infected mice had lower histopathology scores (histological activity index, HAI) early during infection compared to WT infected mice (Figure 3.2C). Even after the bacterial burden in *omasA* infected mice caught up to WT levels (Day 127) the HAI scores trended lower in *omasA* mutant infected mice compared to WT infected mice (p=0.07). The lower histopathology of the *omasA*

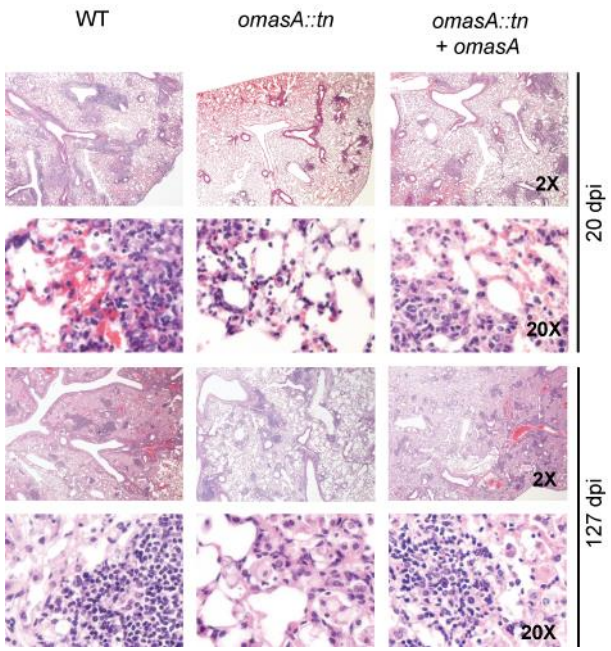

Figure 3.3. Mice infected with the *omasA* mutant have reduced histopathology compared to WT infected mice. A single lung lobe from was fixed and H&E stained for histology. Shown are representative images captured under 2X and 20X magnification from 20 and 127 days post infection comparing WT (MBTB178), *omasA::tn* (MBTB319), and *omasA::tn* + *omasA* (MBTB320).

mutant infected mice may help account for their longer survival time in comparison to WT infected mice.

A structural prediction for OmasA reveals similarities with Mce-associated proteins

A β-lactamase reporter fusion was previously used to identify OmasA as a protein with a C-terminal domain exported to the extracytoplasmic space (Figure 3.1A) (McCann *et al.*, 2011). A single transmembrane domain, as predicted by TMHMM (Krogh *et al.*, 2001), maps close to the N-terminus of the protein. Thus, Rv0199 is a relatively small (24kDa) predicted integral membrane protein with the majority of the protein (155 of 219 total amino acids) located on the cell wall side of the membrane (McCann *et al.*, 2011). To confirm the exported nature of OmasA, subcellular fractions of a *Mycobacterium smegmatis* strain engineered to express a C-terminal HA tagged OmasA were prepared for immunoblot analysis. As expected, OmasA-HA primarily localized to the membrane and cell wall fractions of *M. smegmatis*, with a smaller fraction of OmasA-HA being detected in the soluble fraction, which includes cytoplasmic material (Figure 3.1C).

Consistent with a prior bioinformatics analysis (Casali & Riley, 2007), ClustalW2 (Goujon *et al.*, 2010; Thompson *et al.*, 2002) revealed the C-terminal region of OmasA to have a low level of sequence identity (~10-25%) with Mce-associated (Mas) proteins of actinomycetes. Mas proteins are uncharacterized proteins found downstream of *mce* operons (Casali & Riley, 2007). However, the *omasA* gene is not linked to a *mce* operon, leading us to call it an orphaned *mas* gene (*omas*). ClustalW2 reveals 10-25% identity between any two Mas proteins, which is similar to the low homology shared between OmasA and Mas proteins (Figure 3.4). To gain more insight into potential functional domains of the OmasA protein, we used Phyre 2 (protein homology/analogy recognition engine version 2.0), an online 3D structural prediction program

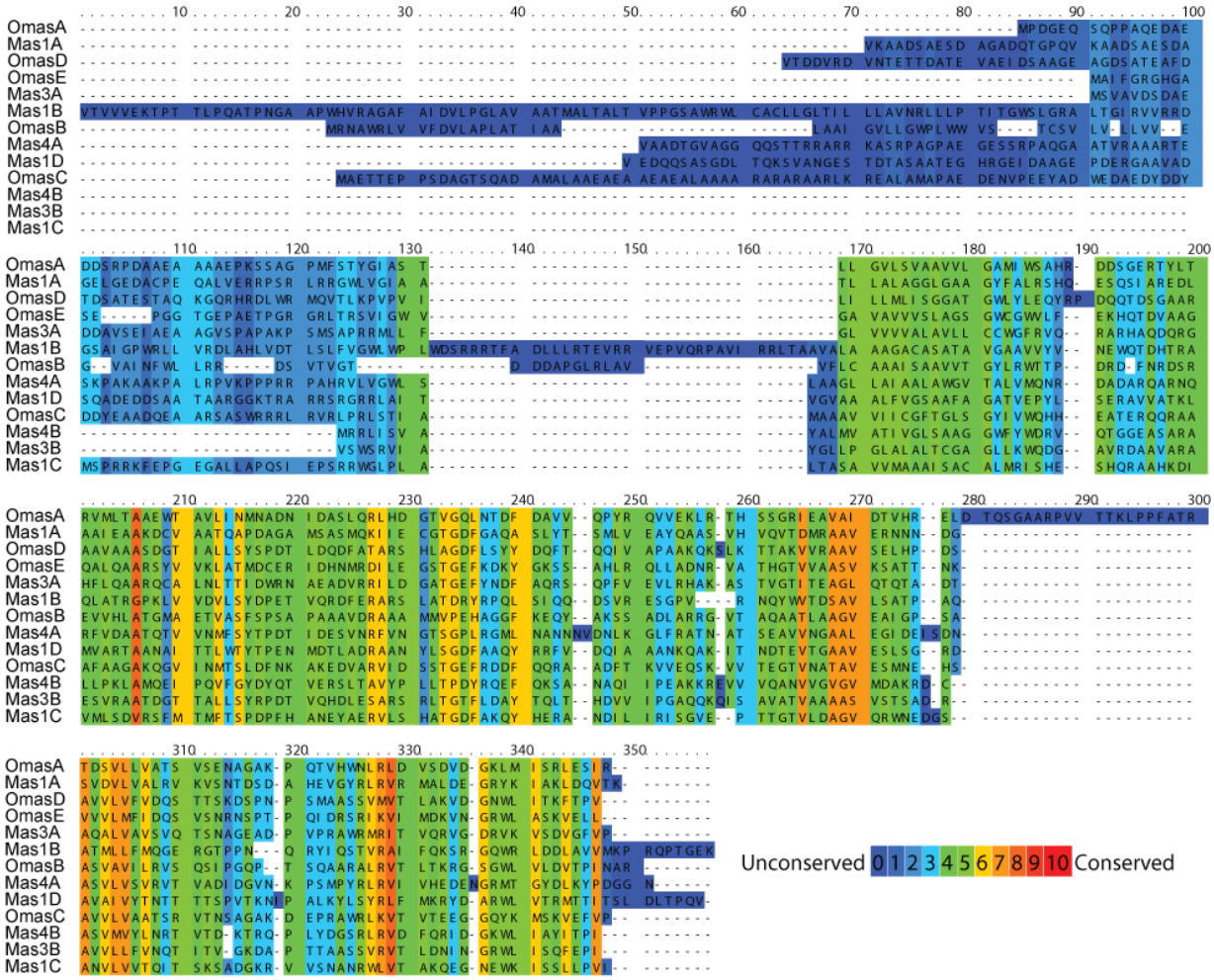

Figure 3.4. **Alignment of all Mas proteins**. Alignment between all predicted *M. tuberculosis* Mas and Omas proteins in PRALINE (Bawono & Heringa, 2014). Colors represent amino acid conservation with blue representing no/low conservation and red indicating high level conservation.

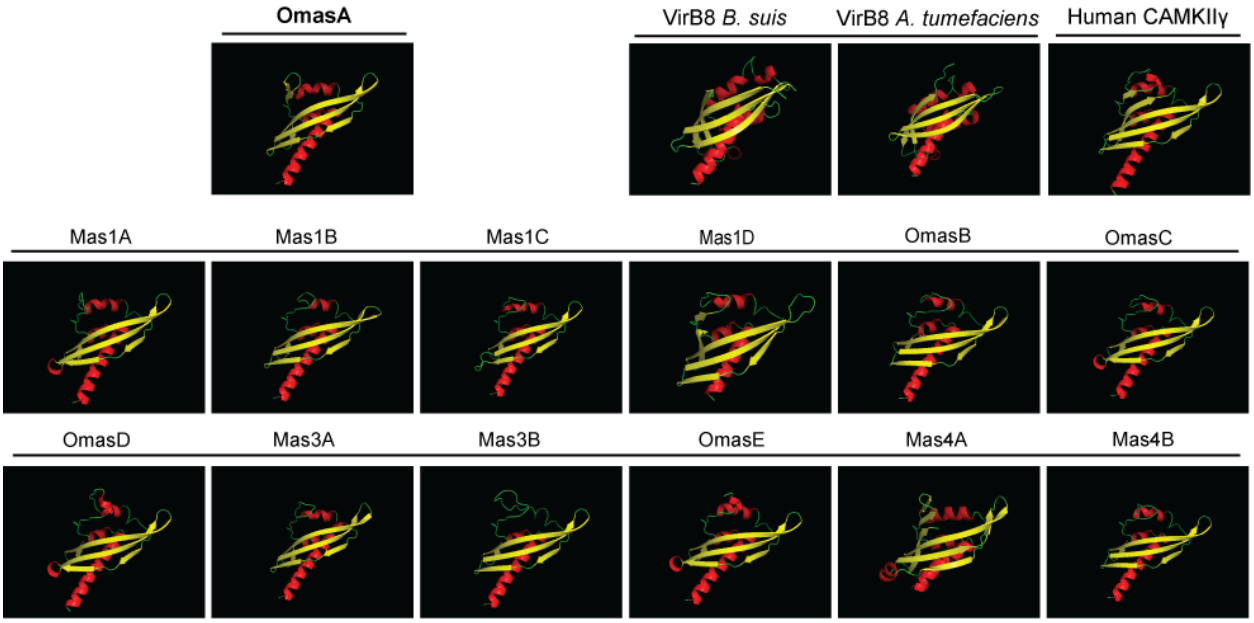

Figure 3.5. Phyre 2, an online structural prediction program, predicts with high confidence that all Mas and Omas proteins form a NTF2-like fold. Ribbon diagrams shown represent Phyre 2 (Kelley & Sternberg, 2009) predicted structures of OmasA, Mas and Omas proteins colored by secondary structure in Pymol. Ribbon diagrams representing the solved crystal structures of the human CAMKIIγ (Rellos *et al.*, 2010) as well as VirB8 from *Brucella suis* and *Agrobacterium tumefaciens* are shown for comparison (Smith *et al.*, 2012; Bailey *et al.*, 2006; Terradot *et al.*, 2005). Alpha helices are colored in red, Beta-strands in yellow, and turns in green.

(Kelley & Sternberg, 2009). Phyre 2 modeled the C-terminal domain of OmasA (aa 69-212) with 96% confidence to the crystal structure of the protein-protein interaction domain of the eukaryotic protein CAMKII (Rellos *et al.*, 2010), predicting OmasA to form a fold that is characteristic of NTF2 family proteins (Figure 3.5) (Chaillan-Huntington *et al.*, 2001). Phyre 2 also detected structural similarity between the C-terminal domain of OmasA and NTF2-like domains in some bacterial proteins. Notable matches were to the structures of VirB8 proteins from *Brucella suis* and *Agrobacterium tumefaciens* (Figure 3.1D) (Smith *et al.*, 2012; Bailey *et al.*, 2006; Terradot *et al.*, 2005). Like OmasA, VirB8 is a small protein, 26 kDa, with an N-terminal transmembrane domain, and the majority of the protein localized to the periplasm. Additionally, like OmasA, in *B. abortus* VirB8 plays an important role during infection of both mice and macrophages (den Hartigh *et al.*, 2008). VirB8 is a component of the type IV secretion system, a large multi-protein transporter, and it is important to both the stability and function of the transporter complex (Kumar *et al.*, 2000; Sivanesan & Baron, 2011; den Hartigh *et al.*, 2008). To determine whether these structural predictions for OmasA are shared with Mas family proteins, we used Phyre 2 to predict the structure of all *M. tuberculosis* Mas proteins. Strikingly, like OmasA, all *M. tuberculosis* Mas family proteins had high confidence structural alignments to NTF2 domain containing proteins, including VirB8 (Figure 3.5). Given the similarity between the structural predictions of OmasA, Mas and VirB8 proteins, we hypothesized a function of OmasA in Mce transporters, possibly a function analogous to that of VirB8 stabilizing multi-protein transporter complexes.

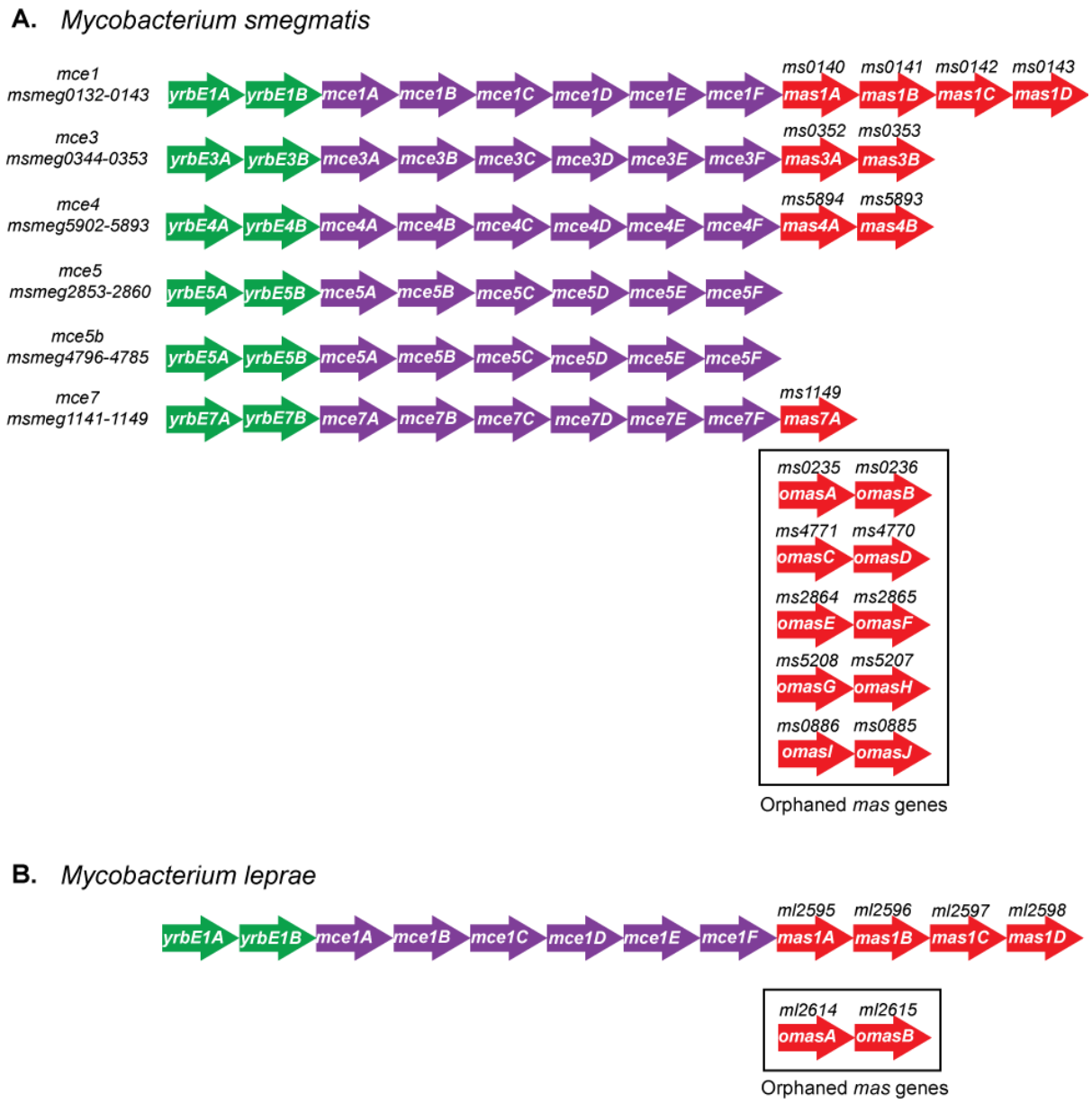

Figure 3.6. Mce operons in *M. smegmatis* and *M. leprae*. A. *M. smegmatis* has six *mce* operons organized as shown. Mce operons are organized by two *yrbE* genes upstream (green), six *mce* genes (purple) and many have *mas* genes (red) downstream. Several genes, including *msmeg0235*, are predicted to encode Mas proteins but are not located in *mce* operons and thus are labeled *orphaned mce-associated* genes (*omas*) (Casali & Riley, 2007). **B.** The *mce* operon structure in *Mycobacterium leprae* which only contains one *mce* operon and two *orphaned mas* genes (Casali & Riley, 2007).

Deletion of omasA in Mycobacterium smegmatis leads to a mce mutant morphology phenotype

With the goal of assigning a function to OmasA, we first explored the potential for OmasA to contribute to Mce transport in *M. smegmatis*. *M. smegmatis* has six *mce* operons with nine *mce-associated* genes (*mas*), and ten *orphaned mce-associated* (*omas*) genes (Casali & Riley, 2007) (Figure 3.6). In *M. smegmatis*, *msmeg0235* is the ortholog of *omasA*, and will be referred to as $omasA_{ms}$. Like $OmasA_{mtb}$, the $OmasA_{ms}$ protein has a predicted transmembrane domain near the N-terminus, and $OmasA_{ms}$ has 55% identity and 76% similarity to $OmasA_{mtb}$ in the C-terminal domain according to BLAST (Altschul *et al.*, 1990). We constructed a deletion mutant of $omasA_{ms}$ and compared phenotypes of the $omasA_{ms}$ mutant to those of *M. smegmatis* mutants lacking *mce4* or all six *M. smegmatis mce* operons (*mce6X*) (Klepp *et al.*, 2012).

Previous studies revealed a rugose morphology for the *mce6X M. smegmatis* mutant growing on Mueller Hinton agar plates containing Congo red (Klepp *et al.*, 2012). Consequently, we tested whether the $omasA_{ms}$ mutant displays a similar morphology. Plates were incubated at 37°C for two days and morphology was assessed by low-magnification microscopy. WT *M. smegmatis* displayed flat, shiny colonies, but *mce4*, *mce6X*, and the $omasA_{ms}$ mutants displayed rugose morphology (Figure 3.7). The rugose phenotype of the $omasA_{ms}$ mutant could be complemented by either expression of $omasA_{mtb}$ or $omasA_{ms}$ from a plasmid (Figure 3.7). While the basis of the *mce* mutant rugose phenotype is not currently understood, the appearance of a similar phenotype for the $omasA_{ms}$ *M. smegmatis* mutant is consistent with a role for OmasA in Mce transporters.

Double mutants $omasA_{ms}$*mce4* and $omasA_{ms}$*mce6X* mutants were also constructed and tested for possible epistatic interactions. Double mutants were spotted and compared to single

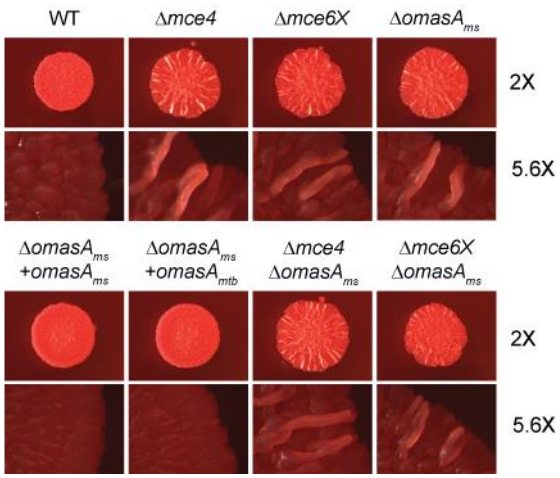

Figure 3.7. The *omasA*$_{ms}$ mutant shares a morphology phenotype with *mce* operon mutants. Two µL spots of culture were plated on Mueller Hinton plates containing glucose and Congo red. The resulting colonies were visualized after 2 days at 2X and 5.6X magnification (Leica M420 macroscope). Results are representative of at least three independent experiments comparing WT +pMV261 (EP1182), *ΔmceA* +pMV261 (EP1204), *Δmce6X* +pMV261 (EP1208), *ΔomasA* +pMV261 (EP1193), *ΔomasA* +*omasA*$_{ms}$ (EP1194), *ΔomasA* +*omasA*$_{mtb}$ (EP1203), *ΔomasAΔmceA* +pMV261 (EP1206), and *ΔomasAΔmce6X* +pMV261 (EP1210). pMV261 is an empty vector, *omasA* expression constructs are cloned in pMV261.

mce4 or *mce6X* mutants (Figure 3.7). If the rugose phenotype of the *omasA*$_{ms}$ mutant is due to the effective loss of Mce transport, the double *omasA*$_{ms}$*mce4* and *omasA*$_{ms}$*mce6X* should look like single *mce4* or *mce6X* mutants. If rugosity of the *omasA*$_{ms}$ mutant is independent of Mce transporter function, an additive effect on rugose morphology from losing both *mce* operons and *omasA*$_{ms}$ may occur. The double mutant phenotype was indistinguishable from that of the single mutants, suggesting that OmasA functions in the Mce transporter pathway.

OmasA is required for cholesterol utilization

The Mce4 transporter has a demonstrated function in cholesterol import. Mycobacterial mutants lacking the *mce4* operon are defective in cholesterol uptake and growth on cholesterol as a sole carbon source (Klepp *et al.*, 2012; Pandey & Sassetti, 2008). To test whether OmasA is specifically important to Mce4 function, we tested the *omasA*$_{ms}$ mutant for its ability to grow in liquid media with cholesterol as a sole carbon source. *M. smegmatis* strains were grown in minimal media supplemented with standard glucose and glycerol carbon sources or cholesterol as the sole carbon source. Cell number and metabolic activity in media containing different carbon sources was followed by resazurin reduction over time, as previously described (Hayden *et al.*, 2013). Resazurin is a blue dye that converts to a pink fluorescent compound when reduced by metabolically active cells. In glucose and glycerol containing media, the resazurin reduction observed over time for the *omasA*$_{ms}$ mutant was equivalent to that exhibited by WT *M. smegmatis* (Figure 3.8A). However, in media with cholesterol as a sole carbon source, resazurin reduction was observed with WT *M. smegmatis* but the *omasA*$_{ms}$ mutant showed very little to no resazurin reduction. Strikingly, the behavior of the *omasA*$_{ms}$ mutant in cholesterol media was equivalent to that of the *mce4 M. smegmatis* mutant (Figure 3.8B). The cholesterol growth defect

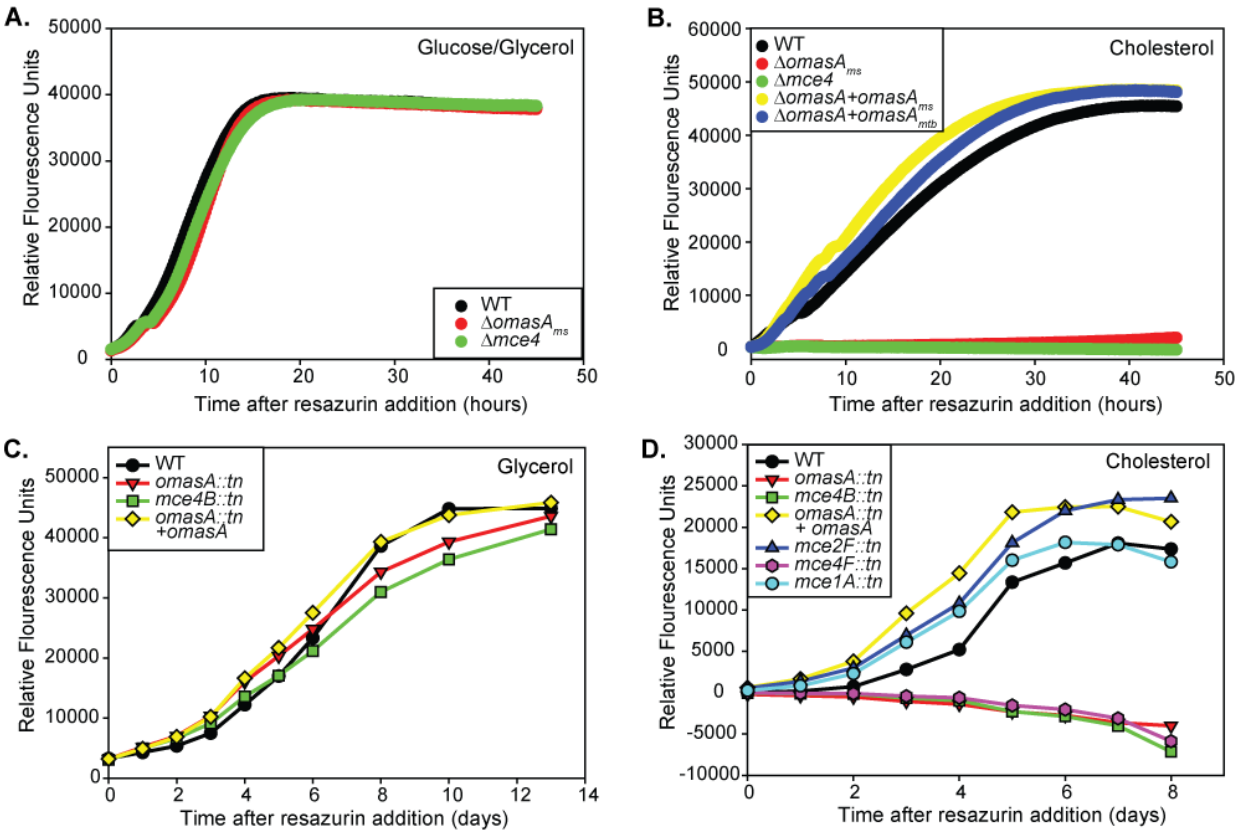

Figure 3.8. OmasA is required for *M. smegmatis* and *M. tuberculosis* to utilize cholesterol. A. 10^4 colony forming units (cfu) of *M. smegmatis* strains were added to M9 glucose/glycerol and growth was monitored by resazurin conversion over time. **B.** 10^4 cfu of *M. smegmatis* were added to minimal M9 media plus cholesterol, and growth was monitored by resazurin conversion over time. Relative fluorescence unit measurements in cholesterol media are reported after subtraction of the minimal signal from no carbon source. 10^4 cfu of *M. tuberculosis* were added to minimal Sauton's media supplemented with **C.** glycerol or **D.** cholesterol, and growth was monitored by resazurin conversion over time. Relative fluorescence unit measurements in cholesterol media are reported after subtraction of the minimal signal from no carbon source. Results are representative of at least three independent experiments. *M. smegmatis* strains: WT +pMV261 (EP1182), *Δmce4* +pMV261 (EP1204), *ΔomasA* +pMV261 (EP1193), *ΔomasA* +$omasA_{ms}$ (EP1194), and *ΔomasA* +$omasA_{mtb}$ (EP1203). *M. tuberculosis* strains: WT (MBTB178), *omasA::tn* (MBTB319), *omasA::tn + omasA* (MBTB320), *mce2F::tn* (MBTB156), *mce1A::tn* (MBTB204), *mce4B::tn* (MBTB329), *mce4F::tn* (MBTB288).

of the *omasA*$_{ms}$ mutant could be fully complemented by expression of either *omasA*$_{ms}$ or *omasA*$_{mtb}$ from a plasmid (Figure 3.8B).

To determine whether OmasA is also important to Mce4 function in *M. tuberculosis*, we tested the *M. tuberculosis omasA* mutant for a defect in growth with cholesterol as a sole carbon source. Again we used resazurin reduction over time to monitor cell number and metabolic activity. Transposon mutants in several *mce* operons, including the *mce4* operon (insertion mutants in the *mce4B* and *mce4F* genes), *mce1* (*mce1B*), and *mce2* (*mce2F*) were tested in parallel with the *omasA* mutant and complemented strains. The *omasA* mutant reduced resazurin to an equivalent level as WT when grown in glycerol media (Figure 3.8C). However, in cholesterol media, the *omasA* mutant and *mce4* mutants with transposon insertions in *mce4B* or *mce4F* were unable to utilize cholesterol as a sole carbon source (Figure 3.8D). As with the *M. smegmatis* cholesterol growth experiments, the *omasA* and the *mce4* mutants of *M. tuberculosis* exhibited the same level of defect in cholesterol media. The *omasA* mutant was fully complemented by expression of *omasA*$_{mtb}$ in the complemented strain. Transposon mutants interrupting *mce1* and *mce2* operons displayed no defect for utilization of cholesterol (Figure 3.8D), consistent with previous reports (Pandey & Sassetti, 2008; Griffin *et al.*, 2011). These data demonstrate that OmasA is required for cholesterol utilization in both *M. smegmatis* and *M. tuberculosis*.

OmasA is required for cholesterol uptake

The cholesterol growth defects of *omasA* mutants of *M. smegmatis* or *M. tuberculosis* were indistinguishable from those of *mce4* mutants, suggesting a role of OmasA in Mce4 cholesterol import. To directly test whether OmasA is required for cholesterol import, as opposed to playing a role in downstream cholesterol metabolism, we tested the ability of WT *M.*

smegmatis, the *omasA*$_{ms}$ mutant, and complemented strains to import radioactively labeled cholesterol. *M. smegmatis* strains were grown overnight in media with glucose and glycerol and then incubated for two hours in minimal media with C^{14} labeled cholesterol as the sole carbon source. After incubation, cells were washed extensively, and the level of accumulated cholesterol in the cells was quantified. In these experiments, the *mce4* mutant exhibited a two-fold reduction in cholesterol uptake in comparison to WT, consistent with previous reports (Pandey & Sassetti, 2008; Klepp *et al.*, 2012). The *omasA*$_{ms}$ mutant also revealed a defect in cholesterol uptake in comparison to WT, and this defect was equivalent to that observed with the *M. smegmatis mce4* mutant (Figure 3.9A). The cholesterol uptake defect of the *omasA* mutant could be complemented by either *omasA*$_{mtb}$ or *omasA*$_{ms}$ (Figure 3.9A). While both the *mce4* and *omasA*$_{ms}$ mutants exhibited a significant reduction in cholesterol uptake, there remained detectable levels of cell-associated C^{14} cholesterol with both mutants. Previous uptake studies also report residual levels of cholesterol associated with *mce4* mutants, leading to the suggestion that additional cholesterol importers may exist in mycobacteria (Pandey & Sassetti, 2008; Klepp *et al.*, 2012). When we examined the double *omasA*$_{ms}$*mce4* mutant it was no more defective than single *mce4* or *omasA*$_{ms}$ mutants. In fact, the double mutant showed slightly improved cholesterol uptake in comparison to the single *mce4* and *omasA*$_{ms}$ mutations alone (Figure 3.9B). The lack of an additive effect of the *mce4* and *omasA*$_{ms}$ mutations on the cholesterol uptake phenotype is consistent with OmasA functioning in concert with Mce4 to import cholesterol, as opposed to being part of an independent cholesterol uptake pathway.

OmasA stabilizes the Mce1 transport complex

The above studies demonstrate that OmasA is important to Mce4 cholesterol import and utilization. However, a function of OmasA beyond Mce4 seems likely. This is because the role

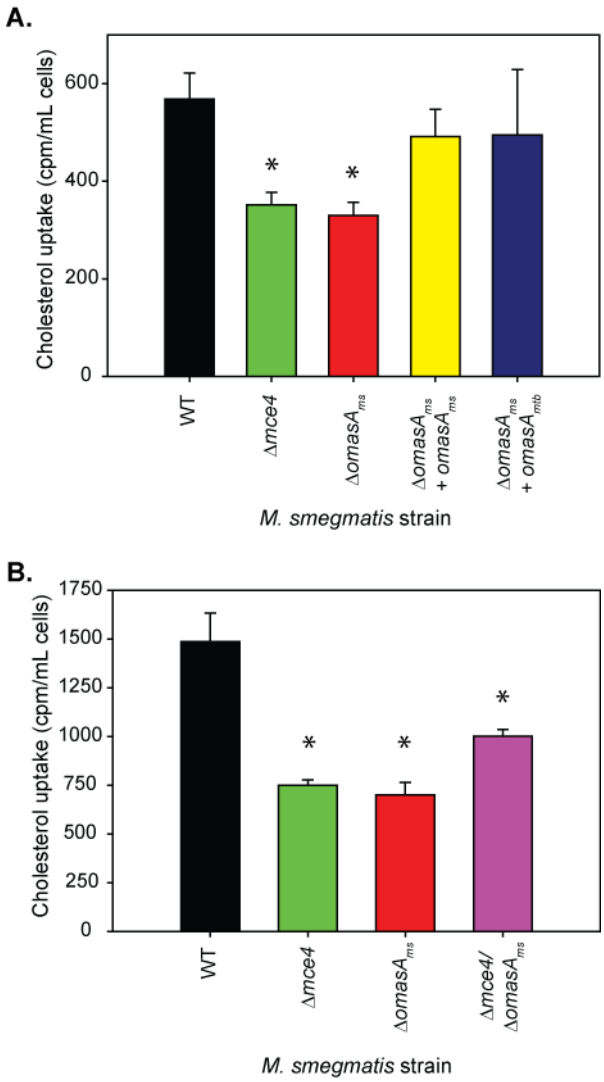

Figure 3.9. OmasA is required for cholesterol uptake. **A.** and **B.** *M. smegmatis* strains were grown overnight in M9 glucose/glycerol, and washed extensively in M9 no carbon source. Cells were incubated with 4-C^{14}-cholesterol for two hours, washed extensively, and cell associated radioactivity levels were measured by scintillation counter. * indicates $p<0.05$ compared to WT. Error bars represent standard deviation. Results are representative of at least three independent experiments. *M. smegmatis* strains: WT +pMV261 (EP1182), *Δmce4* +pMV261 (EP1204), *ΔomasA* +pMV261 (EP1193), *ΔomasA +omasA*$_{ms}$ (EP1194), *ΔomasA +omasA*$_{mtb}$ (EP1203), and *ΔomasAΔmce4*+pMV261 (EP1206).

of OmasA in promoting *M. tuberculosis* growth in resting murine macrophages (McCann *et al.*, 2011) cannot be explained by an effect on Mce4, as there is no obvious role for Mce4 in promoting growth in resting macrophages (Pandey & Sassetti, 2008; Stewart *et al.*, 2005; Rengarajan *et al.*, 2005; McCann *et al.*, 2011). However, because *M. tuberculosis mce1* mutants are reported in several studies to be defective for growth in macrophages (Rengarajan *et al.*, 2005; Stewart *et al.*, 2005; McCann *et al.*, 2011), we hypothesized that OmasA is additionally important to Mce1 transporter function. Due to the predicted structural similarities between OmasA and VirB8, and the role of VirB8 in stabilizing the multi-protein type IV secretion complex (den Hartigh *et al.*, 2008; Sivanesan & Baron, 2011), we further hypothesized that OmasA stabilizes proteins within Mce transporter complexes. Thus, to investigate the potential contribution of OmasA to the Mce1 transporter system and stability of the Mce1 complex we performed immunoblot analysis for three *M. tuberculosis* Mce1 proteins (Mce1A, Mce1E, and Mce1F) in *M. tuberculosis* WT, the *omasA* mutant, and the complemented strain. Mce1A, Mce1E, and Mce1F were localized to the cell wall in *M. tuberculosis* WT and the complemented strain, consistent with previous subcellular localization experiments performed in *M. smegmatis* (Forrellad *et al.*, 2014). However, none of these Mce proteins were detected in the cell wall of the *omasA* mutant (Figure 3.10A). Further, Mce1A, Mce1E, and Mce1F were undetectable in the whole cell lysate of the *omasA* mutant, demonstrating that they were not merely mislocalized in the *omasA* mutant. Mce1 protein levels were fully restored in the complemented strain. The effect of the *omasA* mutation on Mce proteins was not due to a broad defect on cell wall proteins, as shown by equivalent levels of the exported 19kD lipoprotein in cell wall fractions of all three strains (Figure 3.10A).

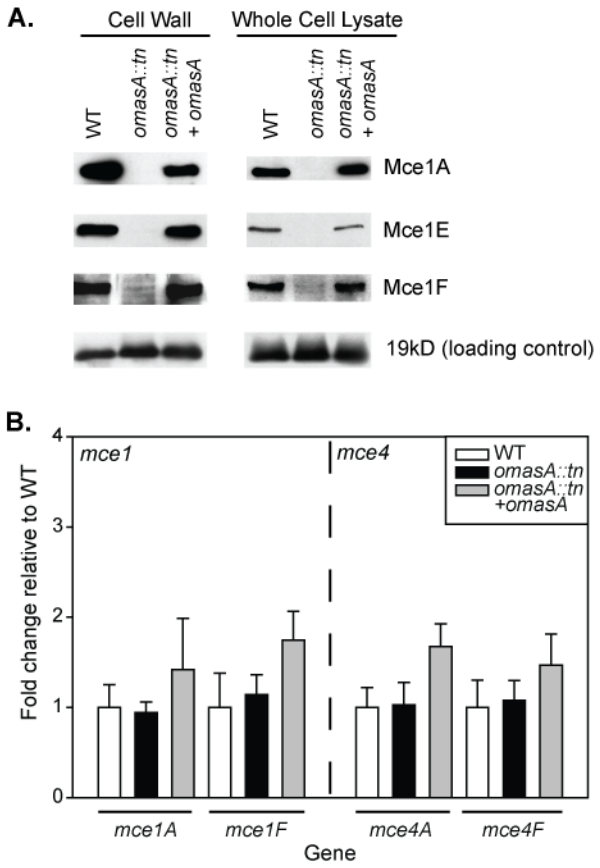

Figure 3.10. Absence of OmasA results in Mce1 protein instability. **A.** *M. tuberculosis* cells were irradiated and lysed by French press to generate whole cell lysates (WCL) and fractionated by differential ultracentrifugation into cell wall fractions. Western blots were performed for Mce1A, Mce1E, Mce1F, and the 19kD lipoprotein. Results are representative of at least three independent replicates. **B.** RNA was collected from *M. tuberculosis* WT, *omasA::tn*, and *omasA::tn +omasA* complemented strains and transcript levels of *mce1A*, *mce1F*, *mce4A*, and *mce4F* were determined by Quantitative Real-Time PCR and normalized to expression of the housekeeping protein *sigA* (Manganelli *et al.*, 1999). Reported are fold change values for each gene relative to expression in WT *M. tuberculosis*. Error bars represent standard deviation. Results are representative of at least three independent biological replicates of WT (MBTB178), *omasA::tn* (MBTB319), and *omasA::tn* + *omasA* (MBTB320).

The immunoblot results are consistent with Mce1 proteins being unstable in the absence of OmasA, however; an alternate explanation is that OmasA is required for the expression of *mce1* genes. To rule out the possibility that the absence of Mce1 proteins in the *omasA* mutant is due to a transcriptional effect, we measured the level of *mce1* transcripts in WT, *omasA* mutant, and complemented strains using Quantitative Real-Time PCR. All three strains harbored equivalent amounts of *mce1A* and *mce1F* transcripts. Thus, the lack of Mce1 proteins in the *omasA* mutant is not a consequence of lower transcript levels. Rather, the striking reduction in Mce1 proteins observed in the *omasA* mutant is consistent with OmasA having a function, similar to that of VirB8, in stabilizing multi-component transporters. We similarly quantified *mce4* transcript levels in the *omasA* mutant and again observed equivalent levels of *mce4A* and *mce4F* transcripts in the *omasA* mutant compared to WT and complemented strains (Figure 3.10B).

Discussion

The goal of this work was to extend our previous identification of OmasA as an exported protein of unknown function with a role in promoting growth in macrophages. Here, we tested the significance of OmasA in a low-dose aerosol model of murine infection and set about assigning a function to the protein. While we detected low level similarity between OmasA and Mas proteins, the homology is limited. Further, the *omasA* gene is not linked to a *mce* operon and there are no prior studies assigning function to any Mas protein. By demonstrating both a role for OmasA during murine infection and a role in Mce1 and Mce4 transporter systems, our results provide the first direct evidence of any Mas protein being required for *in vivo* infection or contributing to Mce transporter function. The lack of assayable *in vitro* phenotypes for the Mce2

and Mce3 systems prevented us from testing a role for OmasA in these other transporters. Thus, it remains a possibility that OmasA is also involved in additional Mce transporter systems.

When tested in mice, the *omasA* mutant exhibited reduced bacterial burden during the growth-*in-vivo* phase of infection (first 3 weeks), which is consistent with the role of OmasA in promoting *M. tuberculosis* growth in macrophages (McCann *et al.*, 2011). The *omasA* mutant infected animals also exhibited reduced pathology and had an increased survival time. Interestingly, later in infection the organ burden of the mutant was no different than WT (as seen in independent experiments) indicating that the basis of the attenuated phenotype of the *omasA* mutant is complex and not simply a consequence of fewer bacteria present throughout infection. Because of the connection we made between OmasA and Mce systems, we compared the macrophage and mouse phenotypes of the *omasA* mutant to infection phenotypes reported for *mce* mutants. In resting macrophages, several studies report intracellular growth defects of *mce1* mutants, like that of the *omasA* mutant (Stewart *et al.*, 2005; Rengarajan *et al.*, 2005; McCann *et al.*, 2011). In mice, a wide variety of models (infection route, mouse strain, etc.) have been used to test *mce* mutants making it difficult to compare studies. However, when tested in mice, mutants in *mce1*, *mce2*, *mce3*, or *mce4* are generally attenuated, as demonstrated by increased survival time, reduced lung pathology, and/or reduced bacterial burden (Marjanovic *et al.*, 2010; Senaratne *et al.*, 2008; Rengarajan *et al.*, 2005; Sassetti & Rubin, 2003; Gioffre *et al.*, 2005), similar to the *omasA* mutant. For an *mce4* mutant specifically, when tested in an intravenous infection model with a 1:1 mixture of WT:*mce4*, the *mce4* mutant is reported to have a persistence defect late in infection (Joshi *et al.*, 2006; Pandey & Sassetti, 2008; Sassetti & Rubin, 2003). However, when tested in an aerosol infection model similar to the one used in our study, a *mce4* mutant exhibits a subtle persistence defect (Senaratne *et al.*, 2008), which may explain why

we did not observe a defect in persistence. Overall, the *omasA* mutant animal phenotypes are consistent with what has been reported for *mce* mutants.

The *omasA* mutant phenotypes we observed on cholesterol-containing media were indistinguishable from *mce4* mutants. These results not only support a role for OmasA in Mce4 transport, but they additionally reveal OmasA to be a new protein required for cholesterol utilization *in vitro*. In a Tn-seq mutagenesis study to identify *M. tuberculosis* genes required for *in vitro* growth on cholesterol, all genes in the *mce4* operon were identified, including *mas4A* and *mas4B*, but *omasA* was not identified (Griffin *et al.*, 2011). Interestingly, *omasA* only barely missed the cutoff for statistical significance in this study (p=0.06), consistent with a role in cholesterol utilization.

It is also interesting to compare our results indicating a role for OmasA in Mce transport pathways to the results of a transposon mutagenesis screen conducted in *mce1* or *mce4* mutant backgrounds (Joshi *et al.*, 2006). In this genetic interaction screen, genes that are members of the same Mce transport pathway or genes in redundant parallel pathways were uncovered. Once again, although *omasA* was not predicted as having genetic interactions with *mce1* or *mce4* in this earlier study, inspection of the supplemental data revealed the behavior of *omasA* mutations in *mce1* and *mce4* backgrounds to be consistent with *omasA* being part of these Mce pathways (Joshi *et al.*, 2006).

Given the many *mce*-linked *mas* genes (eight) and unlinked *omas* (five) genes in *M. tuberculosis*, our finding that deletion of *omasA* yielded phenotypes as dramatic as complete deletion of the *mce4* operon was surprising, as was the discovery that OmasA impacted more than one Mce system. The dramatic phenotypes of the *omasA* mutant raise questions about whether other Mas proteins of *M. tuberculosis* will also have such broad effects. Data from

transposon mutagenesis screens predicts similar phenotypes for mutations in *mas* genes and the adjoining *mce* operons (Rengarajan *et al.*, 2005; Sassetti & Rubin, 2003; Griffin *et al.*, 2011), which supports the idea of *mas* genes functioning with their linked *mce* system. However, the function of *mas* genes may not extend to unlinked *mce* loci. For example, the *mce1*-associated *mas* genes (*mas1A-D*) are not predicted to be required for growth on cholesterol like *mce4* mutants (Griffin *et al.*, 2011). Additionally, it remains unclear whether all orphaned Mas proteins are required for multiple Mce transporters or whether they are even involved in Mce transport at all. Individual *mas* and *omas* mutants will need to be constructed and characterized in order to determine if the dramatic role of OmasA in Mce function is unique or representative of the overall importance of all Mas family members.

The *Mycobacterium leprae* genome is highly reduced in comparison to other mycobacterial species and is thought to have only maintained a minimal set of genes required for its intracellular lifestyle (Moran, 2002; Singh & Cole, 2011). Interestingly, *omasA* and the downstream *omasB* are only two *omas* genes conserved in the *M. leprae* genome, which contains a single *mce* operon, *mce1* (Figure 3.6). Conservation of *omasA* in *M. leprae* supports the importance of *omasA* in intracellular growth and virulence. Additionally, the conserved presence and arrangement of *omasA* and *omasB* suggests that the corresponding proteins may function together. Like *omasA* and *omasB*, Mas family proteins are usually encoded in pairs (Figure 3.1B, Figure 3.6) (Casali & Riley, 2007), although the significance of this arrangement is unknown. Future study of OmasB, the protein encoded by *rv0200*, could help to determine whether OmasB also has a broad role in Mce transport like OmasA.

The unexpected structural similarity predicted by Phyre 2 between VirB8 and Mas proteins was helpful for identifying a function for OmasA. VirB8 is an essential component of

bacterial type IV secretion systems, helping to form the core complex of the transport machinery, providing stability to many proteins within the complex and potentially anchoring it to the cytoplasmic membrane (Paschos *et al.*, 2006; Fronzes *et al.*, 2009; Kumar *et al.*, 2000; Baron, 2006). In the absence of VirB8, many proteins within the type IV secretion apparatus become destabilized and degraded (den Hartigh *et al.*, 2008; Sivanesan & Baron, 2011). Similarly, in the absence of OmasA all three of the Mce1 proteins we monitored by immunoblot analysis were degraded. These results suggest that OmasA, and Mas proteins in general, may play analogous roles to VirB8 in the formation and stabilization of the core Mce transport complex, resulting in destabilization of Mce proteins in their absence. Interestingly, VirB8 also plays a role in substrate transport during type IV secretion (Cascales & Christie, 2004), which raises the possibility that Mas proteins may also have an additional role in Mce substrate movement. Due to structural and functional analogies between OmasA and VirB8, we propose a model wherein OmasA interacts with Mce proteins, potentially driving Mce complex formation, and ultimately providing stability to Mce proteins within the complex (Figure 3.11). Because there are thirteen Mas family proteins in *M. tuberculosis* and only four Mce transporters, we predict that each transporter may be stabilized by multiple Mas family members.

Although Mce transporters are of clear importance to *M. tuberculosis* virulence and a core component of the *M. tuberculosis* genome (Gioffre *et al.*, 2005; Shimono *et al.*, 2003; Marjanovic *et al.*, 2010; Lima *et al.*, 2007; Sassetti & Rubin, 2003; Rengarajan *et al.*, 2005; Stewart *et al.*, 2005; McCann *et al.*, 2011; Senaratne *et al.*, 2008; Pandey & Sassetti, 2008), there has yet to be a systematic genetic or biochemical analysis of the individual Mce transporter proteins in terms of their contribution to virulence or their function in the transport mechanism. Mce transporter components are assigned potential functions by analogy to classic ABC

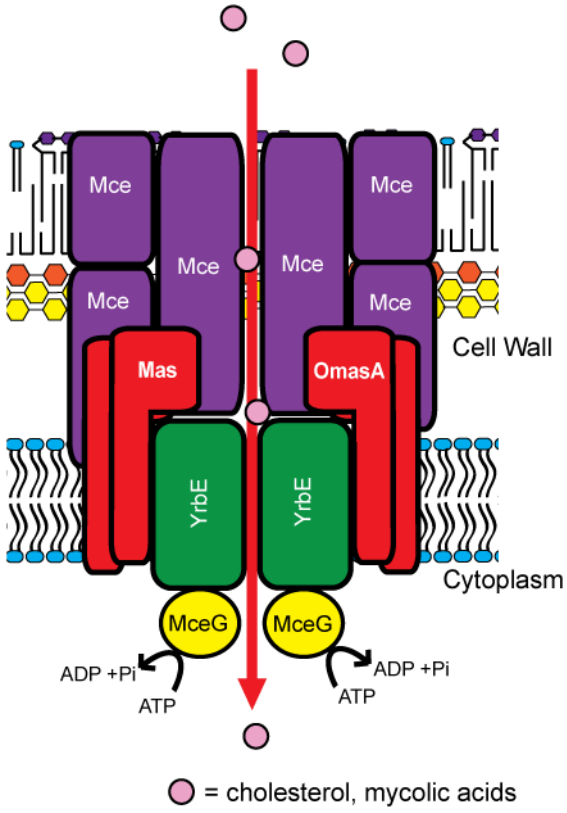

Figure 3.11. OmasA is an integral membrane protein that is important to Mce transporter stability and function. OmasA and other Mas proteins, shown in red, are embedded in the inner membrane by an N-terminal transmembrane domain with the majority of the protein being localized on the cell wall side of the membrane. Mce permease proteins (YrbE), shown in green, are multi-membrane spanning proteins localized to the inner membrane. Some Mce proteins contain predicted TM domains; however, localization from this and other studies (Klepp *et al.*, 2012) suggests that Mce proteins are located within the cell wall (shown in purple). MceG, shown in yellow, is the cytoplasmic Mkl family ATPase predicted to be responsible for ATP-hydrolysis that powers the transport of substrates, shown in pink, through the complex.

transporters (ex. ATPase, permease, or solute binding proteins) (Casali & Riley, 2007). However, Mce transporters are distinguished from ABC transporters in the multitude of individual proteins predicted to be involved: two YrbE permeases, six predicted Mce solute binding proteins, and a shared ATPase MceG. The function of all of these individual transporter components requires validation. Because Mas proteins share no obvious ABC transporter counterpart their function has been an even bigger mystery. The results of this study provide an essential framework for studying the role of Mas family proteins in the assembly and stabilization of Mce transporter systems.

Experimental Procedures

Bacterial strains and plasmids. In this study, we used the bacterial strains listed in Table 3.1 and plasmids as listed in Table 3.2. The *M. tuberculosis omasA* (*rv0199*) mutant was generated in a previous transposon mutagenesis study performed in a *M. tuberculosis* β-lactamase (*ΔblaC*) background (McCann *et al.*, 2011; Flores *et al.*, 2005). The *omasA::tn* mutant has a hygromycin resistant Tn*'bla*$_{TEM-1}$ transposon inserted in the *omasA* coding sequence at amino acid position 74 and it expresses an exported OmasA-'BlaTEM-1 fusion protein. The *omasA::tn* mutant (*omasA::tn, ΔblaC*) used in this study (MBTB319) additionally carries the empty pMV261.kan plasmid. For mutant characterization, *omasA::tn* was compared to strain MBTB178 (*omasA*^{WT}, *ΔblaC,* pJES137, pMV261.kan). Plasmid pJES137 is an integrating hygromycin resistant plasmid that expresses *'blaTEM-1*. MBTB178 is referred to as WT in the text. The *M. tuberculosis* complemented strain (*omasA::tn, ΔblaC,* pJES178) expresses *omasA* from the *hsp60* promoter of the kanamycin resistant plasmid pJES178 (McCann *et al.*, 2011). This series

of *omasA::tn* (MBTB319), *omasA*^{WT} (MBTB178) and complemented (MBTB320) strains are all hygromycin and kanamycin resistant to enable growth in identical media conditions.

Bacterial growth. M. tuberculosis strains were grown in Middlebrook 7H9 broth (Difco) supplemented with 1X albumin dextrose saline (ADS), 0.5% glycerol and either 0.025% Tween 80 (Tw) or 0.025% tyloxapol (Ty). *M. smegmatis* strains were grown in Middlebrook 7H9 broth (Difco) supplemented with 0.2% glucose, 0.5% glycerol, and either 0.05% Tween 80 (Tw) or 0.05% Tyloxapol (Ty). Medium was supplemented with 20µg mL^{-1} kanamycin or 50µg mL^{-1} hygromycin as needed for mycobacterial cultures. *E. coli* strains were grown in Luria-Bertani medium (Fisher) supplemented as necessary with 40µg mL^{-1} kanamycin.

Mouse experiments. Female C57BL/6 mice aged 7-10 weeks were infected with ~200 cfu of *M. tuberculosis* by aerosol using a Madison chamber (Mechanical Engineering Workshop, Madison, WI), and bacterial burden was determined, as previously described (Kurtz *et al.*, 2006). Groups of four mice per strain were sacrificed, organs homogenized, and diluted and plated to determine bacterial burden at various times after infection. The lower right lobe of the lungs was inflated and fixed in 10% formalin for histology.

Histopathology. Inflammation was determined in 5 µm sections following hematoxylin and eosin (H&E) staining. Paraffin embedded sections were set and cut to reveal the maximum longitudinal visualization of the intrapulmonary main axial airway. Histopathology was evaluated and scored by an experienced blinded reviewer (I.C.A.) on a scale of 0 (absent) to 3 (severe), as previously described (McElvania Tekippe *et al.*, 2010; Allen *et al.*, 2013; Allen *et al.*, 2009). The parameters assessed included overall leukocyte infiltration, perivascular and peribroncheolar cuffing, extravasation, and the estimated percent of lung area involved with

inflammation. Each individual parameter was scored and averaged to generate the histology score.

Mutant construction. M. smegmatis mutants were constructed by recombineering, as previously described (van Kessel & Hatfull, 2008; van Kessel & Hatfull, 2007). Briefly, upstream and downstream flanks were PCR amplified and cloned into pMP614 (kind gift from Martin Pavelka), which was then linearized to produce the final recombineering fragment, carrying a hygromycin resistance marker flanked by DNA sequences upstream and downstream of *msmeg0235*. Parental strains carrying a kanamycin marked plasmid expressing a recombinase, pJV53, (van Kessel & Hatfull, 2007; van Kessel & Hatfull, 2008) were used for recombineering. Following three hour induction of the recombinase with acetamide, electroporation was used to introduce the linear recombineering fragment. Allelic exchange recombinants were selected for double resistance to hygromycin and kanamycin. Strains were cured of pJV53 by passaging 3-4 times in the absence of kanamycin. Plasmid cured strains were then transformed with the resolvase expressing pMP854 plasmid (kind gift from Martin Pavelka), to remove the hygromycin marker in the deletion cassette. Hyg^s strains were cured for pMP854 as described above to generate the final unmarked deletion strains. Mutant construction was confirmed by Southern blot (data not shown).

$OmasA_{ms}$ complementation and $OmasA_{mtb}$-HA vector construction. The *msmeg0235* gene (*$omasA_{ms}$*) was PCR amplified by *msmeg0235*_F1 x *msmeg0235*_R1, the *rv0199* gene (*$omasA_{mtb}$*) was PCR amplified by *rv0199HA*_F_MscI x *rv0199HA*_R_HindIII, and PCR fragments were cloned into pCR2.1 (Invitrogen). The resulting plasmids were sequenced to confirm they were error-free. The *$omasA_{ms}$* fragment was digested from pCR2.1 with EcoRI, gel purified, and ligated into EcoRI digested pMV261.kan (Stover *et al.*, 1991). The *$omasA_{mtb}$*

fragment was digested from pCR2.1 with MscI and HindIII, gel purified, and ligated into MscI/HindIII digested JSC77 (Glickman *et al.*, 2000), containing an in-frame C-terminal HA tag. Primer sequences are provided (Table 3.3).

Transformation. M. smegmatis strains were transformed by electroporation, as previously described (Snapper *et al.*, 1990).

Morphology. Congo red assays were performed, as previously described (Klepp *et al.*, 2012). Mueller Hinton agar plates were supplemented with 0.2% glucose and 100µg mL^{-1} Congo red (Sigma). Colony morphology was analyzed by plating 2 µL spots of OD_{600} 1.0 *M. smegmatis* strains. Plates were incubated at 37°C for two days and visualized using a low-magnification Leica M420 macroscope with 2X and 5.6X magnification.

Cholesterol Growth Assays for M. tuberculosis. A cholesterol stock solution was prepared by solubilizing cholesterol in ethanol and tyloxapol, as follows. A 1:1 solution of 200 proof Ethanol:Tyloxapol (Sigma) was prepared, filtered, and heated to 50°C. 200mg mL^{-1} cholesterol was dissolved in 3:1 chloroform:methanol, and added dropwise to the 50°C tyloxapol solution until reaching 20% final volume. Sauton's media was prepared and pH adjusted to 7.4: 1L dH_2O, 4g DL asparagine, 2g sodium citrate, 0.5g K_2HPO_4, 0.5g $MgSO_4$-$7H_2O$, 0.05g ferric ammonium citrate, 0.025% Tyloxapol, and supplemented with either 6% glycerol or 0.5 mM cholesterol from stock solution. *M. tuberculosis* strains were diluted to 10^5 cfu mL^{-1} in Sauton's +Ty and 10^4 cfu were aliquoted into 96 well plates with Sauton's supplemented with glycerol or cholesterol, incubated shaking at 37°C for seven days, then resazurin (Sigma) was added to a final concentration 0.0125 mg mL^{-1}. Resazurin conversion was followed using fluorescence and was monitored daily by a Tecan Infinite 200 Pro at hv=544 nm excitation and hv=590 nm emission.

Cholesterol Growth Assays for M. smegmatis. A cholesterol stock solution was prepared by solubilizing cholesterol in cyclodextrin, as previously described (Klein *et al.*, 1995). Briefly, 1g methyl-ß-cyclodextrin (C4555 Sigma) was dissolved in 11mL PBS (0.09g mL^{-1}) and heated to 80°C with continuous stirring. 30 mg cholesterol (Sigma) was dissolved in 400µL 2:1 isopropanol/chloroform. The cholesterol solution was added to the cyclodextrin in 50µL aliquots, stirring continuously. The solution was cooled slowly, filtered for sterility, and kept at room temperature. M9 minimal media was prepared as follows: 1L dH_2O, 12.8g Na_2HPO_4, 3g KH_2PO_4, 0.5g NaCl, 1g NH_4Cl, 25 µL 1M $CaCl_2$, 500µL 1M $MgSO_4$, and 2.5 mL 10% Tyloxapol (Ty, Sigma), and supplemented with 0.2% glucose and 0.5% glycerol or 0.5mM cholesterol from stock solution. *M. smegmatis* strains were grown to OD_{600} 1.0 in M9 supplemented with 0.2% glucose and 0.5% glycerol + 0.05% Ty. Strains were washed in M9 +Ty three times by pelleting cells at 1,900 x *g* for 10 minutes at 4°C, and diluted to 10^5 cfu mL^{-1} in M9 +Ty, and 10^4 cfu were plated into 96 well plates with M9 containing glycerol or cholesterol. Plates were incubated shaking at 37°C overnight, after which resazurin (Sigma) was added to a final concentration 0.0125 mg mL^{-1}. Florescence was monitored every 10 minutes by a Spectramax M2 using hv=544 nm excitation and hv=590 nm emission.

Cholesterol uptake. Cholesterol uptake experiments were performed, similar to previously reported (Klepp *et al.*, 2012). *M. smegmatis* strains were grown to OD_{600} 1.0 in M9 supplemented with 0.2% glucose and 0.5% glycerol + 0.05% Ty. Strains were washed in M9 + Ty three times by pelleting cells at 1,900 x *g* for 10 minutes at 4°C, and then equalized to OD_{600} 0.5 in M9 + Ty, and incubated with 0.04µCi 4-C^{14} cholesterol (Perkin Elmer NEC018050UC) for 2 hours at 37°C. After incubation, cells were pelleted and washed three times with M9 + Ty, and cell associated radioactivity was measured by scintillation counter.

Subcellular fractionation and Western blotting. M. tuberculosis cells were pelleted by centrifugation (1,900 x *g*) and sterilized by irradiation (JL Shephard Mark I 137Cs irradiator, Department of Radiobiology, University of North Carolina at Chapel Hill). After sterilization, *M. tuberculosis* cells were removed from BSL-3 containment. *M. smegmatis* cells were simply pelleted by centrifugation for 10 minutes at 1,900 x *g*. Subcellular fractionation was then performed, as previously described (Gibbons *et al.*, 2007). Briefly, cells were resuspended in PBS containing protease inhibitors, lysed in a French pressure cell, and unlysed cells were removed by centrifugation (1,900 x *g*). The clarified whole cell lysates (WCL) were subjected to differential ultracentrifugation, 27,000 x *g* for 30 minutes to pellet the cell wall (CW), 100,000 x *g* for 2 hours to pellet the membrane (MEM), and remaining soluble (SOL) fraction containing the cytoplasm. Equal protein amounts, as determined by Bicinchonic acid assay (Pierce), for all fractions and strains were separated by SDS-PAGE and transferred to nitrocellulose membranes. Proteins were detected using the following antibodies: Mce1 antibodies (a gift from Christopher Sassetti, University of Massachusetts Medical School; (Feltcher *et al.*, 2015)): anti-Mce1A (1:10,000), anti-Mce1E/Lprk (1:5,000), anti-Mce1F (1:10,000), anti-19kD (1:20,000) (a gift from Douglas Young, Imperial College, United Kingdom), and anti-HA (1:25,000) (Covance). Anti-mouse and Anti-rabbit IgG conjugated HRP (Biorad) were used as secondary antibodies, as appropriate. HRP signal was detected using Western Lighting Chemiluminescent detection reagent (Perkin-Elmer).

Quantitative Real-Time PCR. Triplicate *M. tuberculosis* cultures were grown to OD_{600} of 1.0 and pelleted by centrifugation for 10 minutes at 1,900 x *g*, and qRT-PCR was performed. Bacteria were lysed by 3:1 chloroform methanol, mixed with Trizol (Invitrogen), and the upper phase was separated and RNA precipitated overnight in isopropanol. RNA samples were pelleted and

washed in 70% Ethanol, and resuspended in RNase-free H_2O. RNA samples were treated with DNase (Promega), purified (Zymo RNA Clean and Concentrator Kit), and converted to cDNA using iScript cDNA Synthesis Kit (BioRad). Triplicate biological and triplicate technical replicates of cDNA from 40 ng RNA each were used for qRT-PCR using the Sensimix SYBR and Flourescein kit (Bioline). Transcript copy number for each gene was calculated as compared to known concentrations of genomic DNA, and each sample was normalized to housekeeping gene *sigA* transcript levels. Primer sequences are provided (Table 3.3).

Statistics. Statistics were performed in SigmaPlot. Normality testing (Shapiro-Wilk) and equal variance testing was done to determine correct statistical methods. Comparisons passing normality and equal variance with two groups were performed by two-tailed Student's t-test. Comparisons not passing normality with two groups were performed by Mann-Whitney rank sum test. Comparisons passing normality and equal variance with more than two groups used one way analysis of variance (ANOVA), followed by multiple comparisons with the Holm-Sidak method as appropriate. Comparisons not passing normality with more than two groups used Kruskal-Wallis one way analysis of variance on ranks, followed by multiple comparisons with Student-Newman-Keuls. Survival was analyzed by Log-rank test followed by multiple comparisons with the Holm-Sidak method.

Attributions

E.F.P. and M.B. designed experiments and wrote the manuscript. E.F.P conducted all experiments except where noted below. J.R.M. identified OmasA as a virulence factor (McCann *et al.*, 2011) and conducted a preliminary experiment demonstrating that OmasA was important during murine infection. J.T.S. assisted with murine experiments to test virulence of an *omasA*

M. tuberculosis mutant. S.M. performed subcellular fractionation of *M. tuberculosis* and Western blots for Mce proteins. I.C.A. evaluated and scored sections for histopathology. V.G. took representative images of histology slides. J.D.H. assisted with microplate fluorescence readings.

Table 3.1. Bacterial strains used in this study		
***M. tuberculosis* strains**	**Description**	**Source**
PM638	*M. tuberculosis* H37Rv *ΔblaC*	Flores et al., 2005
MBTB126	PM638 carrying integrating vector pJES137 (expressing *'blaTEM-1*). Strain is hyg^R	McCann et al., 2011
MBTB156	PM638, *mce2F::tn* (with a *'blaTEM-1* transposon inserted into *mce2F* at amino acid position 476). Strain is hyg^R	McCann et al., 2011
MBTB178	MBTB126 + pMV261.kan. Strain is hyg^R and kan^R	McCann et al., 2011
MBTB183	PM638, *omasA(rv0199)::tn* (with a '*blaTEM-1* transposon inserted in *rv0199* at amino acid position 74). Strain is hyg^R	McCann et al., 2011
MBTB204	PM638, *mce1A::tn* (with a '*blaTEM-1* transposon inserted in *mce1A* at amino acid position 449). Strain is hyg^R	McCann et al., 2011
MBTB288	PM638, *mce4F::tn* (with a '*blaTEM-1* transposon inserted in *mce4F* at amino acid position 539). Strain is hyg^R	McCann et al., 2011
MBTB319	MBTB183 + pMV261. Strain is hyg^R and kan^R	McCann et al., 2011
MBTB320	MBTB183 + pJES178 (expressing $omasA_{mtb}$). Strain is hyg^R and kan^R	McCann et al., 2011
MBTB329	PM638, *mce4B::tn* (with a '*blaTEM-1* transposon inserted in *mce4B* at amino acid position 350). Strain is hyg^R	McCann et al., 2011

***M. smegmatis* strains**	**Description**	**Source**
mc^2155	*M. smegmatis* wild-type (WT)	Snapper et. al, 1990
Δmce4 mutant	*Δmce4*	Klepp et al., 2012
Δmce6X mutant	*Δmce1, Δmce2, Δmce3, Δmce4, Δmce5, Δmce6*	Klepp et al., 2012
EP1178	*ΔomasA, Δmce4*	This study
EP1179	*ΔomasA*	This study
EP1180	*ΔomasA, Δmce6X*	This study
EP1182	mc^2155 + pMV261. Strain is kan^R	This study
EP1193	EP1179 + pMV261. Strain is kan^R	This study
EP1194	EP1179 + pEP139 (expressing $omasA_{ms}$). Strain is kan^R	This study
EP1195	EP1179 + pEP166 (expressing $omasA_{mtb}$HA). Strain is kan^R	This study
EP1203	EP1179 + pJES178 (expressing $omasA_{mtb}$). Strain is kan^R	This study
EP1204	*Δmce4* + pMV261. Strain is kan^R	This study
EP1206	EP1178 + pMV261. Strain is kan^R	This study
EP1208	*Δmce6X* + pMV261. Strain is kan^R	This study
EP1210	EP1180 + pMV261. Strain is kan^R	This study

Table 3.1. Bacterial Strains Used in this Study

Table 3.2. Plasmids used in this study			
Plasmid	**Antibiotic resistance**	**Notes**	**Source**
pMV261	kan	Multicopy mycobacterial vector with *hsp60* promoter	Stover et al., 1991
pJSC77	kan	Multicopy mycobacterial vector, HA tag cloned into pMV261	Glickman et al., 2000
pJV53	kan	Multicopy vector for recombineering, phage recombinase genes expressed from acetamide inducible promoter	van Kessel & Hatfull, 2008; van Kessel & Hatfull, 2007
pMP614	hyg	*res-hyg-res* vector for constructing recombineering vectors	Kind gift from Martin Pavelka
pMP854	kan	Resolvase plasmid to remove *hyg* resistance marker	Kind gift from Martin Pavelka
pJES137	hyg	Integrating mycobacterial vector with consitituitively expressed *'blaTEM-1*	McCann et al., 2011
pJES178	kan	P_{hsp60}-*omasA*$_{mtb}$ cloned into pMV261	McCann et al., 2011
pEP139	kan	P_{hsp60}-*omasA*$_{ms}$ cloned into pMV261	This study
pEP166	kan	P_{hsp60}-*omasA*$_{mtb}$ cloned into pJSC77	This study

Table 3.2. Plasmids Used in this Study

Table 3.3. Primers used in this study		
Gene	**Forward primer**	**Reverse primer**
sigA	GAGATCGGCCAGGTCTACGGCGTG	CTGACATGGGGGCCCGCTACGTTG
mce1A	GTCTTCGCGCTCGTAGTTG	ACGTGTGACCTCCGAGATG
mce1F	TCGTCTCCGTAGTCGCAATC	ACATTGGCCGTCGGATACAG
mce4A	ATCCGCAGTGCTGACTTATC	TGCCACGGTATTTGACCTTG
mce4F	ATCACCGTCATCACCCTGAG	CGTTGGCGTTCTTGTACAGG
msmeg0235	GAGAATTCCATGCCTGACGCGCCCAGCAAG	AGGAATTCGGTTCCTCATCGGATGGGCTCCAGAC
rv0199HA	AAGCTTGTCGAATCGACTCCAACCGGGAGATCA	TTGGCCATGCCTGACGGGGAGCAGAGC

Table 3.3. Primers used in this study.

REFERENCES

Allen, I.C., E. McElvania-TeKippe, J.E. Wilson, J.D. Lich, J.C. Arthur, J.T. Sullivan, M. Braunstein & J.P. Ting, (2013) Characterization of NLRP12 during the in vivo host immune response to *Klebsiella pneumoniae* and *Mycobacterium tuberculosis*. *PLoS One* **8**: e60842.

Allen, I.C., M.A. Scull, C.B. Moore, E.K. Holl, E. McElvania-TeKippe, D.J. Taxman, E.H. Guthrie, R.J. Pickles & J.P. Ting, (2009) The NLRP3 inflammasome mediates in vivo innate immunity to influenza A virus through recognition of viral RNA. *Immunity* **30**: 556-565.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers & D.J. Lipman, (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Bailey, S., D. Ward, R. Middleton, J.G. Grossmann & P.C. Zambryski, (2006) *Agrobacterium tumefaciens* VirB8 structure reveals potential protein-protein interaction sites. *Proc Natl Acad Sci U S A* **103**: 2582-2587.

Baron, C., (2006) VirB8: a conserved type IV secretion system assembly factor and drug target. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **84**: 890-899.

Bawono, P. & J. Heringa, (2014) PRALINE: a versatile multiple sequence alignment toolkit. *Methods Mol Biol* **1079**: 245-262.

Cantrell, S.A., M.D. Leavell, O. Marjanovic, A.T. Iavarone, J.A. Leary & L.W. Riley, (2013) Free mycolic acid accumulation in the cell wall of the mce1 operon mutant strain of *Mycobacterium tuberculosis*. *Journal of microbiology* **51**: 619-626.

Casali, N. & L.W. Riley, (2007) A phylogenomic analysis of the Actinomycetales mce operons. *BMC Genomics* **8**: 60.

Cascales, E. & P.J. Christie, (2004) Definition of a bacterial type IV secretion pathway for a DNA substrate. *Science* **304**: 1170-1173.

Chaillan-Huntington, C., P.J. Butler, J.A. Huntington, D. Akin, C. Feldherr & M. Stewart, (2001) NTF2 monomer-dimer equilibrium. *J Mol Biol* **314**: 465-477.

den Hartigh, A.B., H.G. Rolan, M.F. de Jong & R.M. Tsolis, (2008) VirB3 to VirB6 and VirB8 to VirB11, but not VirB7, are essential for mediating persistence of Brucella in the reticuloendothelial system. *J Bacteriol* **190**: 4427-4436.

Feltcher, M.E., H.P. Gunawardena, K.E. Zulauf, S. Malik, J.E. Griffin, C.M. Sassetti, X. Chen & M. Braunstein, (2015) Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall. *Molecular & cellular proteomics : MCP*.

Flores, A.R., L.M. Parsons & M.S. Pavelka, Jr., (2005) Genetic analysis of the beta-lactamases of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* and susceptibility to beta-lactam antibiotics. *Microbiology* **151**: 521-532.

Forrellad, M.A., L.I. Klepp, A. Gioffre, J. Sabio y Garcia, H.R. Morbidoni, M. de la Paz Santangelo, A.A. Cataldi & F. Bigi, (2013) Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* **4**: 3-66.

Forrellad, M.A., M. McNeil, L. Santangelo Mde, F.C. Blanco, E. Garcia, L.I. Klepp, J. Huff, M. Niederweis, M. Jackson & F. Bigi, (2014) Role of the Mce1 transporter in the lipid homeostasis of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **94**: 170-177.

Fronzes, R., P.J. Christie & G. Waksman, (2009) The structural biology of type IV secretion systems. *Nat Rev Microbiol* **7**: 703-714.

Gibbons, H.S., F. Wolschendorf, M. Abshire, M. Niederweis & M. Braunstein, (2007) Identification of two *Mycobacterium smegmatis* lipoproteins exported by a SecA2-dependent pathway. *J Bacteriol* **189**: 5090-5100.

Gioffre, A., E. Infante, D. Aguilar, M.P. Santangelo, L. Klepp, A. Amadio, V. Meikle, I. Etchechoury, M.I. Romano, A. Cataldi, R.P. Hernandez & F. Bigi, (2005) Mutation in mce operons attenuates *Mycobacterium tuberculosis* virulence. *Microbes Infect* **7**: 325-334.

Glickman, M.S., J.S. Cox & W.R. Jacobs, Jr., (2000) A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. *Mol Cell* **5**: 717-727.

Goujon, M., H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern & R. Lopez, (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* **38**: W695-699.

Griffin, J.E., J.D. Gawronski, M.A. Dejesus, T.R. Ioerger, B.J. Akerley & C.M. Sassetti, (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**: e1002251.

Hayden, J.D., L.R. Brown, H.P. Gunawardena, E.F. Perkowski, X. Chen & M. Braunstein, (2013) Reversible acetylation regulates acetate and propionate metabolism in *Mycobacterium smegmatis*. *Microbiology* **159**: 1986-1999.

Hicks, S.W. & J.E. Galan, (2013) Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. *Nat Rev Microbiol* **11**: 316-326.

Isaac, D.T. & R. Isberg, (2014) Master manipulators: an update on *Legionella pneumophila* Icm/Dot translocated substrates and their host targets. *Future Microbiol* **9**: 343-359.

Joshi, S.M., A.K. Pandey, N. Capite, S.M. Fortune, E.J. Rubin & C.M. Sassetti, (2006) Characterization of mycobacterial virulence genes through genetic interaction mapping. *Proc Natl Acad Sci U S A* **103**: 11760-11765.

Kelley, L.A. & M.J. Sternberg, (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**: 363-371.

Klein, U., G. Gimpl & F. Fahrenholz, (1995) Alteration of the myometrial plasma membrane cholesterol content with beta-cyclodextrin modulates the binding affinity of the oxytocin receptor. *Biochemistry* **34**: 13784-13793.

Klepp, L.I., M.A. Forrellad, A.V. Osella, F.C. Blanco, E.J. Stella, M.V. Bianco, L. Santangelo Mde, C. Sassetti, M. Jackson, A.A. Cataldi, F. Bigi & H.R. Morbidoni, (2012) Impact of the deletion of the six mce operons in *Mycobacterium smegmatis*. *Microbes Infect* **14**: 590-599.

Krogh, A., B. Larsson, G. von Heijne & E.L. Sonnhammer, (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.

Kumar, R.B., Y.H. Xie & A. Das, (2000) Subcellular localization of the *Agrobacterium tumefaciens* T-DNA transport pore proteins: VirB8 is essential for the assembly of the transport pore. *Mol Microbiol* **36**: 608-617.

Kurtz, S., K.P. McKinnon, M.S. Runge, J.P. Ting & M. Braunstein, (2006) The SecA2 secretion factor of *Mycobacterium tuberculosis* promotes growth in macrophages and inhibits the host immune response. *Infect Immun*.

Lew, J.M., A. Kapopoulou, L.M. Jones & S.T. Cole, (2011) TubercuList--10 years after. *Tuberculosis (Edinb)* **91**: 1-7.

Ligon, L.S., J.D. Hayden & M. Braunstein, (2012) The ins and outs of *Mycobacterium tuberculosis* protein export. *Tuberculosis (Edinb)* **92**: 121-132.

Lima, P., B. Sidders, L. Morici, R. Reader, R. Senaratne, N. Casali & L.W. Riley, (2007) Enhanced mortality despite control of lung infection in mice aerogenically infected with a *Mycobacterium tuberculosis mce1* operon mutant. *Microbes Infect* **9**: 1285-1290.

Manganelli, R., E. Dubnau, S. Tyagi, F.R. Kramer & I. Smith, (1999) Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Mol Microbiol* **31**: 715-724.

Marjanovic, O., T. Miyata, A. Goodridge, L.V. Kendall & L.W. Riley, (2010) Mce2 operon mutant strain of *Mycobacterium tuberculosis* is attenuated in C57BL/6 mice. *Tuberculosis (Edinb)* **90**: 50-56.

Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S.T. Cole & R. Brosch, (2004) Macro-array and bioinformatic analyses reveal mycobacterial

'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex. *Microbiology* **150**: 483-496.

McCann, J.R., J.A. McDonough, J.T. Sullivan, M.E. Feltcher & M. Braunstein, (2011) Genome-wide identification of *Mycobacterium tuberculosis* exported proteins with roles in intracellular growth. *J Bacteriol* **193**: 854-861.

McCann, J.R., S. Kurtz & M. Braunstein, (2009) Secreted and exported proteins important to *Mycobacterium tuberculosis* pathogenesis. In: Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis. K. Wooldridge (ed). Norfolk, UK: Caister Academic Press, pp. 265-298.

McElvania Tekippe, E., I.C. Allen, P.D. Hulseberg, J.T. Sullivan, J.R. McCann, M. Sandor, M. Braunstein & J.P. Ting, (2010) Granuloma formation and host defense in chronic *Mycobacterium tuberculosis* infection requires PYCARD/ASC but not NLRP3 or caspase-1. *PLoS One* **5**: e12320.

Mohn, W.W., R. van der Geize, G.R. Stewart, S. Okamoto, J. Liu, L. Dijkhuizen & L.D. Eltis, (2008) The actinobacterial mce4 locus encodes a steroid transporter. *J Biol Chem* **283**: 35368-35374.

Moran, N.A., (2002) Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**: 583-586.

Pandey, A.K. & C.M. Sassetti, (2008) Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A* **105**: 4376-4380.

Paschos, A., G. Patey, D. Sivanesan, C. Gao, R. Bayliss, G. Waksman, D. O'Callaghan & C. Baron, (2006) Dimerization and interactions of *Brucella suis* VirB8 with VirB4 and VirB10 are required for its biological activity. *Proc Natl Acad Sci U S A* **103**: 7252-7257.

Rellos, P., A.C. Pike, F.H. Niesen, E. Salah, W.H. Lee, F. von Delft & S. Knapp, (2010) Structure of the CaMKIIdelta/calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS biology* **8**: e1000426.

Rengarajan, J., B.R. Bloom & E.J. Rubin, (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* **102**: 8327-8332.

Rohde, K., R.M. Yates, G.E. Purdy & D.G. Russell, (2007) *Mycobacterium tuberculosis* and the environment within the phagosome. *Immunological reviews* **219**: 37-54.

Sassetti, C.M. & E.J. Rubin, (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**: 12989-12994.

Senaratne, R.H., B. Sidders, P. Sequeira, G. Saunders, K. Dunphy, O. Marjanovic, J.R. Reader, P. Lima, S. Chan, S. Kendall, J. McFadden & L.W. Riley, (2008) *Mycobacterium*

tuberculosis strains disrupted in mce3 and mce4 operons are attenuated in mice. *Journal of medical microbiology* **57**: 164-170.

Shimono, N., L. Morici, N. Casali, S. Cantrell, B. Sidders, S. Ehrt & L.W. Riley, (2003) Hypervirulent mutant of *Mycobacterium tuberculosis* resulting from disruption of the mce1 operon. *Proc Natl Acad Sci U S A* **100**: 15918-15923.

Singh, P. & S.T. Cole, (2011) *Mycobacterium leprae*: genes, pseudogenes and genetic diversity. *Future Microbiol* **6**: 57-71.

Sivanesan, D. & C. Baron, (2011) The dimer interface of *Agrobacterium tumefaciens* VirB8 is important for type IV secretion system function, stability, and association of VirB2 with the core complex. *J Bacteriol* **193**: 2097-2106.

Smith, M.A., M. Coincon, A. Paschos, B. Jolicoeur, P. Lavallee, J. Sygusch & C. Baron, (2012) Identification of the binding site of Brucella VirB8 interaction inhibitors. *Chem Biol* **19**: 1041-1048.

Snapper, S.B., R.E. Melton, S. Mustafa, T. Kieser & W.R. Jacobs, Jr., (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**: 1911-1919.

Stewart, G.R., J. Patel, B.D. Robertson, A. Rae & D.B. Young, (2005) Mycobacterial mutants with defective control of phagosomal acidification. *PLoS Pathog* **1**: 269-278.

Stover, C.K., V.F. de la Cruz, T.R. Fuerst, J.E. Burlein, L.A. Benson, L.T. Bennett, G.P. Bansal, J.F. Young, M.H. Lee & G.F. Hatfull, (1991) New use of BCG for recombinant vaccines. *Nature* **351**: 456-460.

Terradot, L., R. Bayliss, C. Oomen, G.A. Leonard, C. Baron & G. Waksman, (2005) Structures of two core subunits of the bacterial type IV secretion system, VirB8 from *Brucella suis* and ComB10 from *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **102**: 4596-4601.

Thompson, J.D., T.J. Gibson & D.G. Higgins, (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **Chapter 2**: Unit 2 3.

van Kessel, J.C. & G.F. Hatfull, (2007) Recombineering in *Mycobacterium tuberculosis*. *Nature methods* **4**: 147-152.

van Kessel, J.C. & G.F. Hatfull, (2008) Mycobacterial recombineering. *Methods Mol Biol* **435**: 203-215.

WorldHealthOrganization, (2014) Global Tuberculosis Report 2014.

CHAPTER 4: DISCUSSION

Introduction

Mycobacterium tuberculosis is responsible for the disease tuberculosis, which kills approximately 1.5 million people per year (WorldHealthOrganization, 2014). *M. tuberculosis* produces many proteins that are exported: transported out of the bacterial cytoplasm to the bacterial cell surface and out into the host environment. Exported proteins are located at the host-pathogen interface, in an ideal location to manipulate the host response and allow for intracellular growth, and many exported proteins contribute significantly to virulence (Isaac & Isberg, 2014; Agbor & McCormick, 2011; Ligon *et al.*, 2012). Unfortunately, the approaches to experimentally identify proteins as being exported by bacteria have been limited to study of bacteria growing in laboratory media (*in vitro*). Because *in vitro* laboratory conditions cannot mimic the complexity of the host environment, there are likely critical exported virulence factors that have been missed because they are only exported in the context of infection.

Probing for Bacterial Proteins at the Host-Pathogen Interface

The main objective of this dissertation was to develop a method to identify proteins exported by bacterial pathogens during infection of a host (*in vivo*). To accomplish this goal, we developed a novel method that we refer to as EXIT, EXported In vivo Technology, and applied it to identify *M. tuberculosis* proteins exported during murine infection.

The Comprehensive EXIT Approach

EXIT utilizes the ‘BlaTEM reporter, a reporter originally developed and used in other pathogens (Broome-Smith & Spratt, 1986) and more recently adapted successfully for use in *M. tuberculosis* (McCann *et al.*, 2007; McCann *et al.*, 2011). While the ‘BlaTEM reporter was developed around the same time as the alkaline phosphatase (‘PhoA) reporter, the ‘BlaTEM reporter is often overlooked and it is the ‘PhoA reporter that is routinely used as a genetic reporter of bacterial protein export (Manoil & Beckwith, 1985; Broome-Smith & Spratt, 1986). Like the ‘BlaTEM reporter, the ‘PhoA reporter is only enzymatically active when exported (Manoil & Beckwith, 1985). ‘PhoA reporter constructs are screened with a colorimetric assay, where negative colonies are white and colonies expressing exported ‘PhoA constructs are blue (Manoil & Beckwith, 1985). However, the ‘BlaTEM reporter is even more powerful than the ‘PhoA reporter because it can be used as a selectable marker. For EXIT it was critical that we used a reporter that could be selected for *in vivo*. Only with a selectable reporter were we able to select for and pool bacteria actively producing exported fusions during *in vivo* β-lactam treatment, and use next-generation sequencing to identify exported fusion proteins.

We built a multi-copy plasmid carrying the ‘BlaTEM reporter, constructed a comprehensive library of fusions between *M. tuberculosis* genomic DNA and the ‘BlaTEM reporter, and transformed these plasmids into *M. tuberculosis*. The EXIT library was planned and built on a scale suitable for comprehensive genome-wide analysis (Appendix I), containing $5x10^6$ *M. tuberculosis* clones that included at least one in frame fusion for 99% of *M. tuberculosis* genes. While none of the library construction steps were novel, genetic reporters have rarely been used on such a truly genome-wide scale.

The comprehensive EXIT library was the foundation for genome-wide identification of *in vivo* exported proteins. We infected mice with the EXIT library, and conducted *in vivo* β-lactam selection for clones producing *in vivo* exported proteins. Plasmids from clones recovered after *in vivo* β-lactam selection were analyzed by next-generation sequencing to identify the fusion proteins that had promoted survival and *in vivo* β-lactam resistance. EXIT identified 593 proteins exported by *M. tuberculosis* during infection, including 100 proteins that were never before demonstrated to be exported by any experimental method, 32 proteins that have no *in silico* predicted export signal, and 9 proteins that lack both a predicted export signal and have never been previously identified as exported *in vitro*. Additionally, EXIT identified 6 intergenic unannotated regions that seem to encode entirely new exported proteins.

Many previous studies have applied genetic reporters to study exported proteins. The two most ambitious studies prior to EXIT were conducted by Lewenza *et al.* in *Pseudomonas aeruginosa* (Lewenza *et al.*, 2005) and McCann *et al.* in *M. tuberculosis* (McCann *et al.*, 2011). Each of these studies identified a large number of exported proteins, 310 in *P. aeruginosa* and 111 in *M. tuberculosis*, and represented significant advances to prior smaller scale identifications. To our knowledge, the identification of 593 proteins as exported in EXIT represents the largest number of exported proteins identified by a genetic reporter, to date, as well as the first ever attempt at genome-wide identification of proteins exported *in vivo*.

Although we used EXIT *in vivo*, the more general approach could also be useful for identifying proteins exported during *in vitro* growth. Libraries of ‘BlaTEM fusions can be plated on β-lactam containing agar to select for colonies producing exported fusions *in vitro*, and the colonies that survive and grow after β-lactam selection can then be pooled and next-generation sequencing used to identify *in vitro* exported fusion sites. Because of the selective nature of the

‘BlaTEM reporter this methodology represents a significant advance from traditional ‘PhoA based screens where individual reporter positive clones would be identified from screening a large number of colonies, and then isolated, grown, and sequenced individually (Manoil *et al.*, 1990).

Prior to the development of EXIT there were very few attempts to study bacterial protein export during infection. Of the few studies that had been performed previously, they focused on directly testing export of a single protein previously shown to be exported *in vitro* or predicted to be exported due to genomic proximity to a secretion system (Danelishvili *et al.*, 2010; Pechous *et al.*, 2013; Danelishvili *et al.*, 2014; Ge *et al.*, 2009; Broms *et al.*, 2012). Therefore, the EXIT methodology represents a fundamental advance in the ability to study protein export during infection, and is the first available method to identify *in vivo* exported proteins on a genome-wide scale.

Genetic reporters such as ‘PhoA, ‘BlaTEM, and GFP are not only useful for identifying proteins as being exported but also for providing evidence to support a specific orientation in the membrane (topology). Therefore, information on the location of exported fusion sites in membrane proteins can be used to identify exported domains and bolster *in silico* topology predictions (McCann *et al.*, 2011; Manoil, 1991; Rapp *et al.*, 2004). The EXIT methodology also represents a significant advance from traditional methods of protein topology determination, which traditionally consist of individual construction and testing of C-terminal fusions of membrane proteins engineered in frame with a reporter (Manoil, 1991; Daley *et al.*, 2005). Because of the laborious process of constructing individual in-frame reporter fusions, to our knowledge there is only one example of an attempt at a genome-wide topology study in bacteria (Daley *et al.*, 2005). Towards the goal of better understanding integral membrane protein

topology, Daley *et al.* constructed a library of 1,166 protein fusions to ‘PhoA or GFP reporters and identified the orientation (periplasmic or cytoplasmic) of the C-terminus of 502 membrane proteins in *E. coli* (Daley *et al.*, 2005). EXIT identified 1,293 exported fusions in 337 integral membrane proteins (defined as containing a predicted TM domain but no predicted signal peptide). On average each membrane protein was identified by four fusions spread throughout the length of the protein. In this way EXIT provided a more complete picture of the exported domains of each integral membrane protein than could be generated by C-terminal orientation information alone. Thus, the EXIT methodology is also valuable for its ability to comprehensively determine the topology of membrane proteins, which could vastly increase our knowledge about the 3-dimensional structure of integral membrane proteins.

Identification of *in vivo* Exported Proteins

Statistical modeling was used to identify a population of highly abundant clones which produced *in vivo* exported fusion proteins. EXIT identified 593 proteins as exported during infection in the spleen, and 282 proteins predicted to be exported during infection in the lungs. The vast majority of proteins identified as exported in the lungs were also exported in the spleen (97%), suggesting that *M. tuberculosis* uses a similar strategy to survive in both organs. A large bottleneck in infection greatly limited the number of clones reaching the lungs resulting in incomplete testing of the library in that organ. Therefore, we expect that many of the 311 proteins only identified in the spleen are also exported in the lungs. The eight proteins only identified by EXIT in the lungs and not found as highly abundant in the spleens (including four PE/PPE family proteins) are a very interesting group of proteins for further study. These proteins may reflect organ-specific virulence requirements and could assist in the understanding of how *M. tuberculosis* is able to infect and cause disease in a wide variety of organ systems.

Unpredicted and Unannotated Exported proteins

The 593 proteins exported by *M. tuberculosis* during infection included 32 proteins that had no annotated export signal (transmembrane domain or signal peptide), and 6 intergenic unannotated regions that appeared to encode new exported proteins. Three of the proteins with no annotated export signal and no known function were chosen for follow-up study, and all three proteins were shown to localize to the *M. tuberculosis* cell wall, validating the new proteins identified by EXIT as real exported proteins. An additional protein chosen for follow-up based on a pilot experiment was Rv0990c, a small virulence factor with no predicted export signal and unknown function whose expression is known to be highly induced during infection (Talaat *et al.*, 2004; Abomoelak *et al.*, 2011). Rv0990c was validated as exported in *M. smegmatis* (data not shown). However, Rv0990c was not as highly abundant as many of the other proteins we identified in our large scale experiments, and it did not meet our stringent statistics and is, therefore, not included on the list of 593 EXIT identified proteins. We expect there are other proteins, like Rv0990c, that are exported *in vivo* but were not identified by the stringent statistical analysis we employed in EXIT.

A subset of the 32 proteins with no *in silico* predicted export signal may have been misannotated, and an upstream start site may produce a translated protein with an export signal. In the study described above, Lewenza *et al.* identified 14 proteins with no *in silico* predicted export signal of which 3 proteins possessed signal peptides when translation was initiated at an alternate upstream start site (Lewenza *et al.*, 2005). In light of this, we examine the 32 proteins for potential alternate start sites, and found that the start site was likely misannotated for Rv3035 such that if translation began at the upstream start site the protein would possess an N-terminal transmembrane domain. Mistakes in genome annotation also appear to explain the six intergenic

regions that are predicted to encode exported ORFs in EXIT but do not contain any annotated genes. However, the remaining 31 exported proteins with no predicted export signal do not appear to be explained by annotation mistakes. The four proteins validated as exported were exogenously expressed using the annotated start site and were still exported, suggesting additional reasons that these proteins are not predicted as exported by *in silico* programs. Some proteins could have been missed simply because their export signal deviates enough from standard that it is not recognized by *in silico* prediction algorithms. Or, perhaps they do not possess conventional export signals at all, and may be examples of proteins exported in an unconventional manner. Further study of these proteins may uncover additional specialized export pathways in *M. tuberculosis.* Identification of these new proteins as exported provides important information that could direct future studies of their function in *M. tuberculosis* cellular physiology and virulence.

Topology of membrane proteins

Integral membrane proteins are notoriously difficult to purify and study by biochemical means, thus membrane proteins are highly underrepresented in comparison to soluble proteins in crystallography databases. The lack of 3-dimensional structural data limits functional analyses of these proteins. In Chapter 2, we highlighted specific instances where EXIT provided clarity on the topology for polytopic membrane proteins of particular interest in *M. tuberculosis* virulence and drug resistance. While we chose only a few examples for presenting topology models, the EXIT dataset could be mined to provide topology information for all 337 EXIT identified transmembrane proteins. Information gleaned from individual exported fusion sites provides a valuable resource for future functional studies of membrane proteins, and the exported fusion site database generated by EXIT will be a valuable resource to the *M. tuberculosis* community.

Combining even a small amount of experimental topology data with *in silico* transmembrane prediction programs highly increases the accuracy of the topology model generated (Rapp *et al.*, 2004). In the study described above, Daley *et al.* were able to broaden their results by using experimental C-terminal topology information to constrain the *in silico* transmembrane prediction program TMHMM and generate improved topology models for additional membrane proteins (Daley *et al.*, 2005). One future direction for the EXIT dataset would be to partner with bioinformatics collaborators who could use it to constrain and improve prediction algorithms and provide improved topology models for all *M. tuberculosis* membrane proteins.

In vivo induced export

Bacterial pathogens encounter and must respond to a variety of conditions during infection, and 25% of the *M. tuberculosis* genome has been identified as transcriptionally induced *in vivo* during infection of cultured macrophages, mice, and humans, compared to *in vitro* growth (Dubnau *et al.*, 2002; Schnappinger *et al.*, 2003; Rohde *et al.*, 2007b; Dubnau *et al.*, 2005; Talaat *et al.*, 2007; Rachman *et al.*, 2006a). The original impetus for identifying proteins exported during infection was the hypothesis that some proteins would only be exported, and thus only identified as exported, *in vivo*. We imagine several situations that could result in a protein exported significantly more *in vivo* than *in vitro*: transcriptional and post-transcriptional regulation of protein abundance or direct regulation of protein export. EXIT provided evidence for 38 proteins being *in vivo* induced exported proteins (i.e. proteins that are exported to higher levels during infection than during *in vitro* growth, regardless of mechanism). Of the 38 *in vivo* induced proteins identified, 14 (37%) are known to be transcriptionally induced during infection (Dubnau *et al.*, 2005; Dubnau *et al.*, 2002; Schnappinger *et al.*, 2003; Rachman *et al.*, 2006b ; Talaat *et al.*, 2007; Rohde *et al.*, 2007a; Rohde *et al.*, 2007b; Sharma *et al.*, 2006; Srivastava *et*

al., 2007; Rodriguez *et al.*, 2013). However, the mechanism of transcriptional control for most of these 14 genes has yet to be identified. Additionally, the mechanism for increased export remains unclear for the 23 proteins that have not been shown to be transcriptionally induced. One future direction of EXIT is to study the mechanism of regulation for these proteins to learn more about how *M. tuberculosis* controls its response to infection.

The 38 *in vivo* induced exported proteins include 21 proteins of unknown function, which represent potentially unidentified and underappreciated virulence factors. The remaining *in vivo* induced exported proteins have predicted or demonstrated functions (regulation, host defense, myco-membrane synthesis, and nutrient acquisition) that can teach us about the pressures facing pathogens in the host environment. In particular, many nutrient transporters identified as *in vivo* induced exported proteins are poorly understood, and future study of these transport systems could help identify nutritional requirements and availability *in vivo*. Improved knowledge of the host environment could be applied to design improved *in vitro* growth media that better mimics *in vivo* growth conditions. Manipulation of *in vitro* media conditions has been vital to the study of many virulence strategies including iron acquisition, resistance to oxidative and nitrosative stress, anaerobic growth, and non-replicating persistence. Generation of an improved *in vitro* growth media could facilitate characterization for many *in vivo* induced exported proteins.

Application of EXIT to new systems

The EXIT methodology is theoretically compatible for use in any bacteria that is either naturally, or can be made genetically, β-lactam sensitive. Therefore, EXIT could be used in the future to study the *in vivo* exported proteome for a variety of pathogens, in diverse models of infection, and could lead to a better understanding of host-pathogen interactions for multiple bacterial diseases. In particular, the EXIT methodology could be easily translated to study

bacterial pathogens that are already treated with β-lactam antibiotics and pathogens that demonstrate significant bacterial expansion during infection. Additionally, the EXIT methodology can be further applied to *M. tuberculosis* to identify proteins exported during the persistence phase of infection *in vivo*. Application of EXIT to studying persistence in *M. tuberculosis*, or study of bacterial infections where the bacteria are not growing exponentially, is theoretically possible because β-lactam antibiotics are bactericidal and can kill non-replicating bacteria. In Chapter 2, EXIT was only applied to studying *M. tuberculosis* during the acute phase of infection, and β-lactam treatment was begun one day after inoculation. In order to identify proteins exported during persistence phase, β-lactam treatment would be delayed until after bacterial expansion has ceased (~21 days post infection). However, adapting EXIT to study of latent non-replicating *M. tuberculosis* during persistence would require additional optimization to ensure adequate selection for β-lactam resistant bacteria.

Characterization of in vivo exported proteins

70% of exported proteins in *M. tuberculosis* have no known function which reflects the unique complexity of the mycobacterial cell envelope, evolutionary divergence of mycobacteria from more extensively studied bacterial pathogens, and unique virulence strategies used by *M. tuberculosis* during infection.

<u>New exported proteins as virulence factors</u>

Proteins identified by EXIT as *in vivo* induced exported proteins are particularly intriguing virulence candidates, especially the 21 *in vivo* induced exported proteins of unknown function. 13 of the 38 *in vivo* induced exported proteins (34%) have demonstrated or predicted roles in virulence (Reddy *et al.*, 2013; Danelishvili *et al.*, 2010; Chuang *et al.*, 2015;

Papavinasasundaram *et al.*, 2005; Rifat *et al.*, 2014; Tischler *et al.*, 2013; Marjanovic *et al.*, 2010; McCann *et al.*, 2011; Gioffre *et al.*, 2005; Senaratne *et al.*, 2008; Dutta *et al.*, 2010; Hu *et al.*, 2010; Sassetti & Rubin, 2003; Stewart *et al.*, 2005; Rengarajan *et al.*, 2005). Most of these predictions are from large scale virulence screens (TraSH, TnSeq, DeADMAn) where the whole genome is tested for genes required for virulence. However, we expect that the list of 25 *in vivo* induced proteins that are not predicted to be important for virulence include additional virulence factors missed by these methods. Large scale virulence screens are highly valuable, but they have limits, and there are many published examples of proteins shown to be individually important to virulence that were not predicted to be virulence factors (Forrellad *et al.*, 2013). In particular, secreted and surface associated proteins are most likely to be trans-complemented when tested in a large pool of mutants (e.g. large virulence screens), because their phenotypes could be rescued by exported proteins of surrounding bacteria resulting in a false-negative prediction. Because *M. tuberculosis* regulates *in vivo* induced exported proteins in a spatial or temporal manner, and these proteins are positioned at the host-pathogen interface, we predict the *in vivo* induced exported proteins will include additional unpredicted virulence factors.

To begin to test this hypothesis we obtained ten mutants from a library of transposon mutants (generated by (Lamichhane *et al.*, 2005; Lamichhane *et al.*, 2003) and available through BEI Resources) in genes that encode *in vivo* induced proteins that had not been predicted by virulence screens as virulence factors or individually studied for roles in *M. tuberculosis* virulence. The first mutant we were able to obtain was in *eccD2*, a gene encoding the predicted membrane channel for the ESX-2 secretion system and highlighted in topology models shown in Chapter 2 (Figure 2.12). ESX-1 and ESX-5 secretion systems are known to be important for *M. tuberculosis* virulence (Bottai *et al.*, 2012; Guinn *et al.*, 2004); however, ESX-2 was not

predicted as important for virulence by large scale virulence screens (Dutta *et al.*, 2010; Rengarajan *et al.*, 2005; Stewart *et al.*, 2005; Sassetti & Rubin, 2003; Lamichhane *et al.*, 2005; Zhang *et al.*, 2013). EccD2 was initially chosen for follow-up because it behaved like an *in vivo* induced exported protein, although it did not end up reaching our stringent statistical cutoff for being an *in vivo* induced exported protein, in the end. Preliminary study determined that mice infected with an *eccD2::tn* mutant survive longer than mice infected with wild type *M. tuberculosis* (data not shown), suggesting that ESX-2 secretion plays an important virulence role during infection.

Nine additional transposon mutants were obtained in genes that encode *in vivo* induced exported proteins. Eight of these mutants are in genes of unknown function. We are currently testing these mutants for intracellular growth defects during infection of murine bone marrow derived macrophages. Preliminary results suggest that some of these *in vivo* induced exported proteins not previously predicted to be virulence factors may be required for *M. tuberculosis* growth in macrophages (data not shown). The function of these proteins remains unknown (for eight) and understudied (for one), and future studies will focus on further characterizing the function and virulence role for these exported proteins.

Construction of a functional genomics database for *M. tuberculosis*

In many regards EXIT represents a new tool for functional genomics. EXIT was a genome-wide attempt to identify exported proteins as way to better understand the function of each individual identified protein. In fact, 70% of the EXIT identified exported proteins have unknown function, and knowledge about their subcellular localization will further their functional characterization. Identification of *in vivo* induced exported proteins provides

additional functional clues, and highlights the 38 *in vivo* induced proteins as particularly interesting candidates for future studies.

In an effort to consolidate a vast wealth of data on export, virulence predictions, and *in vivo* expression data we built a functional genomics database that could become a valuable resource to the *M. tuberculosis* community. We combined information from previous publications studying *in vitro* protein export, large scale virulence studies (e.g. TraSH, TnSeq, DeADMAn), and *in vivo* expression data (e.g. microarray, IVET) (Rosenkrands *et al.*, 2000; Malen *et al.*, 2011; Bell *et al.*, 2012; Gu *et al.*, 2003; Xiong *et al.*, 2005; Mawuenyega *et al.*, 2005; Malen *et al.*, 2007; Gunawardena *et al.*, 2013; Wolfe *et al.*, 2010; Gomez *et al.*, 2000; Braunstein *et al.*, 2000; McDonough *et al.*, 2008; McCann *et al.*, 2011; Dutta *et al.*, 2010; Rengarajan *et al.*, 2005; Stewart *et al.*, 2005; Sassetti & Rubin, 2003; Lamichhane *et al.*, 2005; Zhang *et al.*, 2013; Sassetti *et al.*, 2003; Dubnau *et al.*, 2002; Schnappinger *et al.*, 2003; Rohde *et al.*, 2007b; Dubnau *et al.*, 2005; Talaat *et al.*, 2007; Rachman *et al.*, 2006a). We supplemented this previously published data with data generated by submitting the *M. tuberculosis* proteome to *in silico* prediction programs for export, structure, and function. We first analyzed the entire proteome of *M. tuberculosis* for export signals using online bioinformatics tools to predict transmembrane domains (TMHMM, TMPred), or signal peptides (SignalP). We combined this with previous whole genome analyses of Tat signal peptides and lipoprotein signal peptides (McDonough *et al.*, 2008; Sutcliffe & Harrington, 2004). Because of its usefulness in predicting a function for Rv0199/OmasA (Chapter 3), we analyzed the entire *M. tuberculosis* proteome in Phyre (Kelley & Sternberg, 2009) to generate additional functional predictions, and included these hypothetical function predictions alongside annotated functional information from the H37Rv RefSeq genome (released January 9 2012). In this way, we built a database of

information for each protein coding gene that can be mined to identify and predict the function of exported proteins and shed light on *M. tuberculosis* pathogenesis. A table representing portions of this database is included in Appendix III.

Orphaned Mce-associated protein A (OmasA)

One of the *in vivo* exported proteins identified by EXIT, Rv0199/OmasA, was functionally characterized in Chapter 3. Rv0199 is a 24 kDa protein that is predicted to be an integral membrane protein of unknown function (Krogh *et al.*, 2001). In Chapter 3, we characterized the role of Rv0199 in *M. tuberculosis* virulence using a low dose aerosol model of murine infection, and demonstrated that OmasA plays a crucial role in lipid transport by Mce transporters leading us to rename Rv0199 as OmasA (orphaned Mce-associated protein A).

In mice, the *omasA* mutant displays reduced bacterial burden early during infection, reduced tissue pathology, and delayed mouse mortality (Figure 3.2, Figure 3.3). Unfortunately, discrepancies in the literature surrounding *mce* mutant phenotypes in mice (in particular *mce1* phenotypes) distracts from our ability to directly compare *mce* mutants to the *omasA* mutant. Direct comparison of different *mce* mutants in the same background and same infection model would go far to improve our understanding of the role of Mce transport during infection. Thus, one future direction for the OmasA project could be side-by-side comparison of the *omasA* mutant to *mce* mutants in a murine infection model. Additionally, the interaction or cross-talk between different Mce systems is poorly understood and worth exploring further. The evidence that OmasA impacts multiple *mce* systems warrants side-by-side comparison of the *omasA* mutant and mutants missing multiple *mce* operons. However, given the difficulty in making even single mutants in *M. tuberculosis*, it would not be easy to construct such a strain.

While the *mce* mutant phenotypes in mice are controversial, the picture is much clearer in macrophages. Mutants in *mce1* are defective for intracellular growth; however, *mce4* is not required for growth in macrophages (McCann *et al.*, 2011; Miner *et al.*, 2009; Rengarajan *et al.*, 2005; Stewart *et al.*, 2005). Therefore, we hypothesize that the intracellular growth defect of the *omasA* mutant can be attributed to loss of Mce1 transporter function during growth in macrophages.

Due to structural and functional analogies identified between OmasA and VirB8 (den Hartigh *et al.*, 2008; Sivanesan & Baron, 2011; Paschos *et al.*, 2006; Fronzes *et al.*, 2009; Kumar *et al.*, 2000; Baron, 2006; Kelley & Sternberg, 2009), we proposed a model wherein OmasA interacts with Mce proteins, potentially driving Mce complex formation, and ultimately providing stability to Mce proteins within the complex (Figure 3.11). However, many questions remain: Is the function of OmasA actually in stabilizing the Mce complex? What proteins does OmasA interact with (the channel itself, or perhaps just the Mce proteins in the cell wall)? Is OmasA responsible for assembling the multi-protein complex? And what distinguishes different Mas/Omas proteins from one another?

The question of whether OmasA really functions to stabilize the complex is the most pressing. While we have demonstrated that loss of OmasA results in loss of Mce1 proteins and Mce4 transport function but does not affect transcript levels of *mce* operons, we have not directly shown that the lack of Mce proteins is due to their instability. Because of the structural homology between OmasA and VirB8, we hypothesize that the function of OmasA is stability of the Mce complex. However, we have not ruled out a potential role for OmasA in translation of Mce proteins. If OmasA is directly interacting with Mce proteins and providing a scaffold for stabilizing the Mce complex, we would expect to see an enhanced degradation rate and reduced

half-life of Mce proteins in an *omasA* mutant as compared to WT. Preliminary evidence supports this hypothesis and our current model, and suggests that Mce proteins do have a reduced half-life in an *omasA* mutant as compared to WT *M. smegmatis* (data not shown).

We hypothesize that OmasA directly interacts with the Mce complex, either through interaction with the channel (YrbE proteins), other Mas family proteins, or Mce proteins themselves. If OmasA physically interacts with the Mce complex proteins, we may be able to perform co-immunoprecipitation experiments or bacterial two-hybrid (Karimova *et al.*, 2000) experiments confirming these interactions. Previous attempts to screen OmasA for interacting proteins using the bacterial two-hybrid system were unsuccessful due to an apparent high false positive rate (data not shown). This can potentially be explained because we attempted to use the entire OmasA protein instead of producing OmasA without its N-terminal transmembrane domain (Gueguen *et al.*, 2011). However, this warrants revisiting, especially now that we have candidate interacting proteins to directly test and better available negative controls. Preliminary studies have shown that OmasA-HA appears to localize properly in *M. smegmatis,* and can be successfully isolated by immunoprecipitation (Figure 3.1, data not shown), suggesting that co-immunoprecipitation could be potentially be used to identify proteins directly interacting with OmasA.

Evidence from study of Type IV complexes in *Agrobacterium tumefaciens* suggests that VirB8 is responsible for assembly of the multi-protein complex (Kumar *et al.*, 2000), and we hypothesize that OmasA may perform a similar function. For *A. tumefaciens* the role of VirB8 in complex assembly was shown by localizing VirB9 and VirB10 by immuno-electron microscopy and immuno-fluorescence microscopy in WT and *ΔvirB8* strains. In the presence of VirB8, VirB9 and VirB10 formed clusters, suggesting complex formation; however, VirB9 and VirB10

were more randomly localized in the *ΔvirB8* mutant. A similar approach could be utilized to study Mce complex assembly in mycobacteria.

Mce proteins localize to the cell wall in mycobacteria (Forrellad *et al.*, 2014); however, the method by which they localize there is not fully understood. One possibility is that OmasA could play a role in assembling the Mce complex by directing localization of Mce proteins to the cell wall. Therefore, we could perform subcellular fractionation and determine whether Mce proteins localize differently in the *ΔomasA* mutant than in WT *M. smegmatis*. This would be difficult to study with endogenous proteins because Mce proteins are almost completely undetectable in the *ΔomasA* mutant; however, we now have strains of *M. smegmatis* that conditionally overexpress Mce proteins, such that we can detect robust levels of Mce proteins even in the *ΔomasA* mutant ((Forrellad *et al.*, 2014), data not shown).

Aside from the eight *mas* genes located within *mce* operons, *M. tuberculosis* encodes five potential *orphaned mas* genes located outside of *mce* operons, including *rv0199/omasA* (Casali & Riley, 2007). As well as identifying structural homology between OmasA and VirB8, Phyre 2 (Kelley & Sternberg, 2009) predicted structural similarity between all Mas family proteins and VirB8. Therefore, we hypothesize that all Mas and Omas proteins may perform similar functions to OmasA in stability and/or assembly of Mce transporters. Given that *M. tuberculosis* possesses thirteen Mas family proteins it is perhaps surprising that the loss of a single one of these proteins, OmasA, results in a dramatic reduction in Mce1 protein levels and Mce4 transport function. If Mas family proteins are functionally analogous, why does *M. tuberculosis* produce so many different varieties and why is there no compensation for loss of OmasA? One potential explanation is that each separate system (Mce1-4) requires multiple individualized Mas family

proteins that provide compatible but not functionally redundant roles. Alternatively, OmasA may play a more critical role than most Mas family proteins.

Several avenues could be explored to address these questions. One possibility for why alternate Mas and Omas proteins did not compensate for the loss of OmasA is that they are not produced in the conditions we tested in sufficient amount to make up for the loss of OmasA. Therefore, we overexpressed Rv0199-like Mas and Omas proteins in the *omasA* mutant background of *M. tuberculosis* to test for functional complementation. Preliminary studies of these overexpression strains in the *omasA* mutant have not shown any evidence of compensation for the loss of *omasA* (data not shown). An alternate strategy to address this question may be to generate additional mutants in *mas* and *omas* genes to directly compare their phenotypes. If physical interaction studies with OmasA are successful, it would also be valuable to identify and compare the proteins that interact with different Mas and Omas proteins. Alignments or Phyre structural predictions of Mas and Omas proteins do not suggest any obvious differences (primary amino acid alignments have very low conservation across the board), or any domains possessed by OmasA that are missing from the other proteins. However, if crystal structures of Mas family proteins could be solved, comparisons could reveal unique structural differences required for their individual function. Overall, future studies are necessary to determine whether the function of OmasA is representative of all Mas proteins, and to functionally characterize other proteins within the Mce transport complex. We are just beginning to understand the function of OmasA, and future studies promise to refine our understanding of Mce complex assembly and stability.

Conclusion

Prior to this work new exported proteins could only be identified *in vitro*; however, we have developed a method to identify exported proteins *in vivo* in the context of infection. In its first application, EXIT successfully identified *in vivo* exported proteins in *M. tuberculosis*, including proteins not previously demonstrated as exported *in vitro*, not predicted to be exported, and *in vivo* induced exported proteins. However, identification of exported proteins is a proximal goal. Each of these categories of proteins provides new and exciting avenues to study *M. tuberculosis* virulence and interaction with the host immune system. With the knowledge that exported proteins are critical to *M. tuberculosis* virulence we seek to better understand the roles of individual exported proteins during infection and assign function to these exported proteins. OmasA is just the beginning, the characterization of one *in vivo* exported protein of unknown function with a previously unknown role in virulence. Having completed an extensive project designed to identify *in vivo* exported proteins and constructed a functional genomics database to help identify the most interesting proteins for future study, the focus will now shift to applying the EXIT dataset to uncover novel virulence strategies of *M. tuberculosis* and extend our understanding of the host-pathogen interface.

REFERENCES

Abomoelak, B., S.A. Marcus, S.K. Ward, P.C. Karakousis, H. Steinberg & A.M. Talaat, (2011) Characterization of a novel heat shock protein (Hsp22.5) involved in the pathogenesis of *Mycobacterium tuberculosis*. *J Bacteriol* **193**: 3497-3505.

Agbor, T.A. & B.A. McCormick, (2011) Salmonella effectors: important players modulating host cell function during infection. *Cell Microbiol* **13**: 1858-1869.

Baron, C., (2006) VirB8: a conserved type IV secretion system assembly factor and drug target. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **84**: 890-899.

Bell, C., G.T. Smith, M.J. Sweredoski & S. Hess, (2012) Characterization of the *Mycobacterium tuberculosis* proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. *J Proteome Res* **11**: 119-130.

Bottai, D., M. Di Luca, L. Majlessi, W. Frigui, R. Simeone, F. Sayes, W. Bitter, M.J. Brennan, C. Leclerc, G. Batoni, M. Campa, R. Brosch & S. Esin, (2012) Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. *Mol Microbiol* **83**: 1195-1209.

Braunstein, M., T.I. Griffin, J.I. Kriakov, S.T. Friedman, N.D. Grindley & W.R. Jacobs, Jr., (2000) Identification of genes encoding exported *Mycobacterium tuberculosis* proteins using a Tn552'phoA in vitro transposition system. *J Bacteriol* **182**: 2732-2740.

Broms, J.E., L. Meyer, K. Sun, M. Lavander & A. Sjostedt, (2012) Unique substrates secreted by the type VI secretion system of *Francisella tularensis* during intramacrophage infection. *PLoS One* **7**: e50473.

Broome-Smith, J.K. & B.G. Spratt, (1986) A vector for the construction of translational fusions to TEM beta-lactamase and the analysis of protein export signals and membrane protein topology. *Gene* **49**: 341-349.

Casali, N. & L.W. Riley, (2007) A phylogenomic analysis of the Actinomycetales mce operons. *BMC Genomics* **8**: 60.

Chuang, Y.M., N. Bandyopadhyay, D. Rifat, H. Rubin, J.S. Bader & P.C. Karakousis, (2015) Deficiency of the novel exopolyphosphatase Rv1026/PPX2 leads to metabolic downshift and altered cell wall permeability in *Mycobacterium tuberculosis*. *mBio* **6**: e02428.

Daley, D.O., M. Rapp, E. Granseth, K. Melen, D. Drew & G. von Heijne, (2005) Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* **308**: 1321-1323.

Danelishvili, L., L. Babrak, S.J. Rose, J. Everman & L.E. Bermudez, (2014) Mycobacterium tuberculosis alters the metalloprotease activity of the COP9 signalosome. *mBio* **5**.

Danelishvili, L., Y. Yamazaki, J. Selker & L.E. Bermudez, (2010) Secreted *Mycobacterium tuberculosis* Rv3654c and Rv3655c proteins participate in the suppression of macrophage apoptosis. *PLoS One* **5**: e10474.

den Hartigh, A.B., H.G. Rolan, M.F. de Jong & R.M. Tsolis, (2008) VirB3 to VirB6 and VirB8 to VirB11, but not VirB7, are essential for mediating persistence of Brucella in the reticuloendothelial system. *J Bacteriol* **190**: 4427-4436.

Dubnau, E., J. Chan, V.P. Mohan & I. Smith, (2005) Responses of *Mycobacterium tuberculosis* to growth in the mouse lung. *Infect Immun* **73**: 3754-3757.

Dubnau, E., P. Fontan, R. Manganelli, S. Soares-Appel & I. Smith, (2002) *Mycobacterium tuberculosis* genes induced during infection of human macrophages. *Infect Immun* **70**: 2787-2795.

Dutta, N.K., S. Mehra, P.J. Didier, C.J. Roy, L.A. Doyle, X. Alvarez, M. Ratterree, N.A. Be, G. Lamichhane, S.K. Jain, M.R. Lacey, A.A. Lackner & D. Kaushal, (2010) Genetic requirements for the survival of tubercle bacilli in primates. *J Infect Dis* **201**: 1743-1752.

Forrellad, M.A., L.I. Klepp, A. Gioffre, J. Sabio y Garcia, H.R. Morbidoni, M. de la Paz Santangelo, A.A. Cataldi & F. Bigi, (2013) Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* **4**: 3-66.

Forrellad, M.A., M. McNeil, L. Santangelo Mde, F.C. Blanco, E. Garcia, L.I. Klepp, J. Huff, M. Niederweis, M. Jackson & F. Bigi, (2014) Role of the Mce1 transporter in the lipid homeostasis of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **94**: 170-177.

Fronzes, R., P.J. Christie & G. Waksman, (2009) The structural biology of type IV secretion systems. *Nat Rev Microbiol* **7**: 703-714.

Ge, J., H. Xu, T. Li, Y. Zhou, Z. Zhang, S. Li, L. Liu & F. Shao, (2009) A Legionella type IV effector activates the NF-kappaB pathway by phosphorylating the IkappaB family of inhibitors. *Proc Natl Acad Sci U S A* **106**: 13725-13730.

Gioffre, A., E. Infante, D. Aguilar, M.P. Santangelo, L. Klepp, A. Amadio, V. Meikle, I. Etchechoury, M.I. Romano, A. Cataldi, R.P. Hernandez & F. Bigi, (2005) Mutation in mce operons attenuates *Mycobacterium tuberculosis* virulence. *Microbes Infect* **7**: 325-334.

Gomez, M., S. Johnson & M.L. Gennaro, (2000) Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. *Infect Immun* **68**: 2323-2327.

Gu, S., J. Chen, K.M. Dobos, E.M. Bradbury, J.T. Belisle & X. Chen, (2003) Comprehensive Proteomic Profiling of the Membrane Constituents of a *Mycobacterium tuberculosis* Strain. *Molecular & cellular proteomics : MCP* **2**: 1284-1296.

Gueguen, E., J. Flores-Kim & A.J. Darwin, (2011) The *Yersinia enterocolitica* phage shock proteins B and C can form homodimers and heterodimers in vivo with the possibility of close association between multiple domains. *J Bacteriol* **193**: 5747-5758.

Guinn, K.M., M.J. Hickey, S.K. Mathur, K.L. Zakel, J.E. Grotzke, D.M. Lewinsohn, S. Smith & D.R. Sherman, (2004) Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* **51**: 359-370.

Gunawardena, H.P., M.E. Feltcher, J.A. Wrobel, S. Gu, M. Braunstein & X. Chen, (2013) Comparison of the membrane proteome of virulent *Mycobacterium tuberculosis* and the attenuated *Mycobacterium bovis* BCG vaccine strain by label-free quantitative proteomics. *J Proteome Res* **12**: 5463-5474.

Hu, Y., R. van der Geize, G.S. Besra, S.S. Gurcha, A. Liu, M. Rohde, M. Singh & A. Coates, (2010) 3-Ketosteroid 9alpha-hydroxylase is an essential factor in the pathogenesis of *Mycobacterium tuberculosis*. *Mol Microbiol* **75**: 107-121.

Isaac, D.T. & R. Isberg, (2014) Master manipulators: an update on *Legionella pneumophila* Icm/Dot translocated substrates and their host targets. *Future Microbiol* **9**: 343-359.

Karimova, G., A. Ullmann & D. Ladant, (2000) A bacterial two-hybrid system that exploits a cAMP signaling cascade in *Escherichia coli*. *Methods Enzymol* **328**: 59-73.

Kelley, L.A. & M.J. Sternberg, (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**: 363-371.

Krogh, A., B. Larsson, G. von Heijne & E.L. Sonnhammer, (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.

Kumar, R.B., Y.H. Xie & A. Das, (2000) Subcellular localization of the *Agrobacterium tumefaciens* T-DNA transport pore proteins: VirB8 is essential for the assembly of the transport pore. *Mol Microbiol* **36**: 608-617.

Lamichhane, G., S. Tyagi & W.R. Bishai, (2005) Designer arrays for defined mutant analysis to detect genes essential for survival of *Mycobacterium tuberculosis* in mouse lungs. *Infect Immun* **73**: 2533-2540.

Lamichhane, G., M. Zignol, N.J. Blades, D.E. Geiman, A. Dougherty, J. Grosset, K.W. Broman & W.R. Bishai, (2003) A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **100**: 7213-7218.

Lewenza, S., J.L. Gardy, F.S. Brinkman & R.E. Hancock, (2005) Genome-wide identification of *Pseudomonas aeruginosa* exported proteins using a consensus computational strategy combined with a laboratory-based PhoA fusion screen. *Genome Res* **15**: 321-329.

Ligon, L.S., J.D. Hayden & M. Braunstein, (2012) The ins and outs of *Mycobacterium tuberculosis* protein export. *Tuberculosis (Edinb)* **92**: 121-132.

Malen, H., F.S. Berven, K.E. Fladmark & H.G. Wiker, (2007) Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics* **7**: 1702-1718.

Malen, H., G.A. De Souza, S. Pathak, T. Softeland & H.G. Wiker, (2011) Comparison of membrane proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra strains. *BMC Microbiol* **11**: 18.

Manoil, C., (1991) Analysis of membrane topology using alkaline phosphatase and β-galactosidase gene fusions. In: Methods in Cell Biology. A.M. Tartakoff (ed). New York: Academic Press, Inc., pp. 61-75.

Manoil, C. & J. Beckwith, (1985) TnphoA: a transposon probe for protein export signals. *Proc Natl Acad Sci U S A* **82**: 8129-8133.

Manoil, C., J.J. Mekalanos & J. Beckwith, (1990) Alkaline phosphatase fusions: sensors of subcellular location. *J Bacteriol* **172**: 515-518.

Marjanovic, O., T. Miyata, A. Goodridge, L.V. Kendall & L.W. Riley, (2010) Mce2 operon mutant strain of *Mycobacterium tuberculosis* is attenuated in C57BL/6 mice. *Tuberculosis (Edinb)* **90**: 50-56.

Mawuenyega, K.G., C.V. Forst, K.M. Dobos, J.T. Belisle, J. Chen, E.M. Bradbury, A.R. Bradbury & X. Chen, (2005) *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. *Mol Biol Cell* **16**: 396-404.

McCann, J.R., J.A. McDonough, M.S. Pavelka & M. Braunstein, (2007) Beta-lactamase can function as a reporter of bacterial protein export during *Mycobacterium tuberculosis* infection of host cells. *Microbiology* **153**: 3350-3359.

McCann, J.R., J.A. McDonough, J.T. Sullivan, M.E. Feltcher & M. Braunstein, (2011) Genome-wide identification of *Mycobacterium tuberculosis* exported proteins with roles in intracellular growth. *J Bacteriol* **193**: 854-861.

McDonough, J.A., J.R. McCann, E.M. Tekippe, J.S. Silverman, N.W. Rigel & M. Braunstein, (2008) Identification of functional Tat signal sequences in *Mycobacterium tuberculosis* proteins. *J Bacteriol* **190**: 6428-6438.

Miner, M.D., J.C. Chang, A.K. Pandey, C.M. Sassetti & D.R. Sherman, (2009) Role of cholesterol in *Mycobacterium tuberculosis* infection. *Indian J Exp Biol* **47**: 407-411.

Papavinasasundaram, K.G., B. Chan, J.H. Chung, M.J. Colston, E.O. Davis & Y. Av-Gay, (2005) Deletion of the *Mycobacterium tuberculosis* pknH gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice. *J Bacteriol* **187**: 5751-5760.

Paschos, A., G. Patey, D. Sivanesan, C. Gao, R. Bayliss, G. Waksman, D. O'Callaghan & C. Baron, (2006) Dimerization and interactions of *Brucella suis* VirB8 with VirB4 and VirB10 are required for its biological activity. *Proc Natl Acad Sci U S A* **103**: 7252-7257.

Pechous, R.D., V. Sivaraman, P.A. Price, N.M. Stasulli & W.E. Goldman, (2013) Early host cell targets of *Yersinia pestis* during primary pneumonic plague. *PLoS Pathog* **9**: e1003679.

Rachman, H., M. Strong, U. Schaible, J. Schuchhardt, K. Hagens, H. Mollenkopf, D. Eisenberg & S.H. Kaufmann, (2006a) *Mycobacterium tuberculosis* gene expression profiling within the context of protein networks. *Microbes Infect* **8**: 747-757.

Rachman, H., M. Strong, T. Ulrichs, L. Grode, J. Schuchhardt, H. Mollenkopf, G.A. Kosmiadi, D. Eisenberg & S.H. Kaufmann, (2006b) Unique transcriptome signature of *Mycobacterium tuberculosis* in pulmonary tuberculosis. *Infect Immun* **74**: 1233-1242.

Rapp, M., D. Drew, D.O. Daley, J. Nilsson, T. Carvalho, K. Melen, J.W. De Gier & G. Von Heijne, (2004) Experimentally based topology models for *E. coli* inner membrane proteins. *Protein science : a publication of the Protein Society* **13**: 937-945.

Reddy, P.V., R.V. Puri, P. Chauhan, R. Kar, A. Rohilla, A. Khera & A.K. Tyagi, (2013) Disruption of mycobactin biosynthesis leads to attenuation of *Mycobacterium tuberculosis* for growth and virulence. *J Infect Dis* **208**: 1255-1265.

Rengarajan, J., B.R. Bloom & E.J. Rubin, (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* **102**: 8327-8332.

Rifat, D., D.A. Belchis & P.C. Karakousis, (2014) senX3-independent contribution of regX3 to *Mycobacterium tuberculosis* virulence. *BMC Microbiol* **14**: 265.

Rodriguez, J.E., A.S. Ramirez, L.P. Salas, C. Helguera-Repetto, J. Gonzalez-y-Merchand, C.Y. Soto & R. Hernandez-Pando, (2013) Transcription of genes involved in sulfolipid and polyacyltrehalose biosynthesis of *Mycobacterium tuberculosis* in experimental latent tuberculosis infection. *PLoS One* **8**: e58378.

Rohde, K., R.M. Yates, G.E. Purdy & D.G. Russell, (2007a) *Mycobacterium tuberculosis* and the environment within the phagosome. *Immunological reviews* **219**: 37-54.

Rohde, K.H., R.B. Abramovitch & D.G. Russell, (2007b) *Mycobacterium tuberculosis* invasion of macrophages: linking bacterial gene expression to environmental cues. *Cell Host Microbe* **2**: 352-364.

Rosenkrands, I., K. Weldingh, S. Jacobsen, C.V. Hansen, W. Florio, I. Gianetri & P. Andersen, (2000) Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. *Electrophoresis* **21**: 935-948.

Sassetti, C.M., D.H. Boyd & E.J. Rubin, (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**: 77-84.

Sassetti, C.M. & E.J. Rubin, (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**: 12989-12994.

Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan & G.K. Schoolnik, (2003) Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *J Exp Med* **198**: 693-704.

Senaratne, R.H., B. Sidders, P. Sequeira, G. Saunders, K. Dunphy, O. Marjanovic, J.R. Reader, P. Lima, S. Chan, S. Kendall, J. McFadden & L.W. Riley, (2008) *Mycobacterium tuberculosis* strains disrupted in mce3 and mce4 operons are attenuated in mice. *Journal of medical microbiology* **57**: 164-170.

Sharma, K., M. Gupta, M. Pathak, N. Gupta, A. Koul, S. Sarangi, R. Baweja & Y. Singh, (2006) Transcriptional control of the mycobacterial embCAB operon by PknH through a regulatory protein, EmbR, in vivo. *J Bacteriol* **188**: 2936-2944.

Sivanesan, D. & C. Baron, (2011) The dimer interface of *Agrobacterium tumefaciens* VirB8 is important for type IV secretion system function, stability, and association of VirB2 with the core complex. *J Bacteriol* **193**: 2097-2106.

Srivastava, V., C. Rouanet, R. Srivastava, B. Ramalingam, C. Locht & B.S. Srivastava, (2007) Macrophage-specific *Mycobacterium tuberculosis* genes: identification by green fluorescent protein and kanamycin resistance selection. *Microbiology* **153**: 659-666.

Stewart, G.R., J. Patel, B.D. Robertson, A. Rae & D.B. Young, (2005) Mycobacterial mutants with defective control of phagosomal acidification. *PLoS Pathog* **1**: 269-278.

Sutcliffe, I.C. & D.J. Harrington, (2004) Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiol Rev* **28**: 645-659.

Talaat, A.M., R. Lyons, S.T. Howard & S.A. Johnston, (2004) The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. *Proc Natl Acad Sci U S A* **101**: 4602-4607.

Talaat, A.M., S.K. Ward, C.W. Wu, E. Rondon, C. Tavano, J.P. Bannantine, R. Lyons & S.A. Johnston, (2007) Mycobacterial bacilli are metabolically active during chronic tuberculosis in murine lungs: insights from genome-wide transcriptional profiling. *J Bacteriol* **189**: 4265-4274.

Tischler, A.D., R.L. Leistikow, M.A. Kirksey, M.I. Voskuil & J.D. McKinney, (2013) *Mycobacterium tuberculosis* requires phosphate-responsive gene regulation to resist host immunity. *Infect Immun* **81**: 317-328.

Wolfe, L.M., S.B. Mahaffey, N.A. Kruh & K.M. Dobos, (2010) Proteomic definition of the cell wall of *Mycobacterium tuberculosis*. *J Proteome Res* **9**: 5816-5826.

WorldHealthOrganization, (2014) Global Tuberculosis Report 2014.

Xiong, Y., M.J. Chalmers, F.P. Gao, T.A. Cross & A.G. Marshall, (2005) Identification of *Mycobacterium tuberculosis* H37Rv integral membrane proteins by one-dimensional gel electrophoresis and liquid chromatography electrospray ionization tandem mass spectrometry. *J Proteome Res* **4**: 855-861.

Zhang, Y.J., M.C. Reddy, T.R. Ioerger, A.C. Rothchild, V. Dartois, B.M. Schuster, A. Trauner, D. Wallis, S. Galaviz, C. Huttenhower, J.C. Sacchettini, S.M. Behar & E.J. Rubin, (2013) Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. *Cell* **155**: 1296-1308.

APPENDIX I: CALCULATIONS TO DETERMINE GENOMIC LIBRARY DENSITY AND MOUSE NUMBERS

Determine size for genomic DNA Library

Original calculations to determine the number of plasmids required for full genome coverage in the EXIT Input library were made using an equation from the Molecular Cloning 2nd Edition (Sambrook *et al.*, 1989). This equation determines the probability that a given DNA sequence is represented within a plasmid or cosmid library. *P* represented the probability of representation, and *f* was the fractional proportion of the genome in a single plasmid or cosmid. *N* was the required number of plasmids or cosmids in the library to reach given probability (*P*) of a single gene of interest being represented *P*.

$$N = \frac{\ln(1-P)}{\ln(1-f)}$$

The equation above was solved separately for three different scenarios where different sizes of genomic DNA fragments were inserted into the library plasmid backbone (pDW31): 500 bp fragments, 1,000 bp fragments, and 5,000 bp fragments. To reach 99% probability of a given gene of interest being represented in the library, the following library sizes (i.e. number of plasmids) were necessary:

$$N = \frac{\ln(1-0.99)}{\ln(1-[\frac{500}{4{,}411{,}532}])} = \frac{-4.61}{-1.13 * 10^{-4}} = 40{,}629\ plasmids$$

$$N = \frac{\ln(1-0.99)}{\ln(1-[\frac{1{,}000}{4{,}411{,}532}])} = 20{,}313$$

$$N = \frac{\ln(1 - 0.99)}{\ln(1 - [\frac{5{,}000}{4{,}411{,}532}])} = 4{,}061$$

This equation is intended to calculate how many cosmids or plasmids are necessary for complete coverage of a genome; however, the EXIT library is additionally constrained in that only plasmids containing a given gene of interest fused in frame with the reporter are capable of producing functional reporter constructs. Therefore, six times more plasmids were required in the EXIT library than would be required in a standard plasmid/cosmid library, and complete coverage of the *M. tuberculosis* genome in frame with the reporter was calculated to require approximately 2.5×10^5 plasmids. By these calculations, the final EXIT library of 5×10^6 plasmids represented 20 (500 bp fragments) to 200 fold (5,000 bp fragments) above minimal requirements for 99% probability of genome coverage.

Another way to evaluate the required size of our library, using a cumulative binomial equation, is more appropriate for this calculation. Cumulative binomial calculations are made far less challenging with the use of an online calculator (e.g. http://stattrek.com/online-calculator/binomial.aspx). In the cumulative binomial calculator a trial (*n*) represented a single plasmid. The number of successes (*x*) was 1, representing a goal of at least one representative plasmid in the library in frame with a given gene of interest. The probability that a single plasmid contained an in frame fusion to a given gene of interest was calculated as:

$$\frac{size\ of\ average\ gene}{genome\ size} * \frac{number\ of\ in\ frame\ fusions}{number\ of\ out\ of\ frame\ and\ opposite\ orientation\ fusions}$$

$$\frac{1{,}006\ bp}{4{,}411{,}532\ bp} * \frac{1}{6} = 3.8 * 10^{-5}$$

Given a sufficiently large *n* value, or number of plasmids containing fusions to the reporter, it was possible to achieve a very high probability of complete genome coverage. For

example, given a library of $2.5x10^5$ plasmids, the probability of an average size gene (1006 bp) represented in frame by at least one fusion, $P(x \geq 1)$, was 99.9%. 99.9% representation for genes as small as 100 bp, or $1/10^{th}$ the average size, required a library of $2x10^6$ plasmids. Therefore, the final EXIT library of $5x10^6$ plasmids theoretically had saturating coverage of genes as small as 55 bp in frame with the reporter.

Determine the number of mice required for comprehensive library testing

In mice, several routes of tuberculosis infection are regularly utilized. A low dose aerosol model is the most similar to natural infection, where humans are thought to be infected by inhalation of as few as one bacterium. However, the low aerosol dosage is not compatible with screening a large library. Further, using aerosol delivery we were unable to increase the pulmonary dosage to a sufficient level to survey the EXIT library in a comprehensive manner (data not shown). We therefore adopted the strategy used for screening *M. tuberculosis* transposon insertion libraries for virulence defects, and infected mice by intravenous infection and focused on comprehensively screening the library in the mouse spleens (Sassetti & Rubin, 2003; Zhang *et al.*, 2013). One day after intravenous infection, approximately 20% of the *M. tuberculosis* inoculum established in the spleen and 1% in the lungs (data not shown), consistent with previous studies (Orme & Gonzalez-Juarrero, 2007).

We calculated the number of animals that would be required in order to comprehensively screen the EXIT library in mouse spleens prior to conducting the large-scale experiments. Using the modified binomial calculation described below we determined that infection of 24 mice with $5x10^6$ bacteria from the EXIT library would result in a 99.5% probability of any individual clone in the library establishing infection in the spleen of at least one mouse. For the modified

binomial calculation *n* was the number of mice/trials, *Q* was the probability of failure on a given mouse/trial, and *P* was the probability of success overall.

$$1 - P = (Q)^n$$

To comprehensively test the EXIT input library in the spleen ($5x10^6$ individual clones) we determined that it would be easiest if each mouse was infected with effectively the entire library, and aimed for infection with $5x10^6$ cfu per mouse. Therefore, an individual library clone injected into a single mouse would have a 20% chance of reaching the spleen (because 20% of the inoculum establishes in the spleen). For this inoculum, there would be a failure rate (i.e. not reaching the spleen) of *Q*=0.8 or 80%. For a lower inoculum dosage, for example $5x10^5$, there would be compounding factors resulting in a higher failure rate, *Q*. An individual library clone would only have a 10% chance of being in the inoculum that is injected into the mouse, and then only 20% chance of reaching the spleen, so only a 2% chance of successfully establishing in the spleen. For this more stringent scenario *Q*=0.98, or 98% chance of failure, and a much higher number of mice would have been required to comprehensively test the EXIT library at this lower initial dosage.

We used the modified binomial equation and solved for the value of *n* to get the required number of mice for the comprehensive testing of the EXIT Input library in the spleen (probability *P*=99.5%) where *Q*=0.8 (24 mice).

$$1 - 0.995 = (0.8)^n$$

Alternatively, comprehensive testing of the EXIT input library in the lung, where only 1% of the inoculum established after intravenous injection (*Q*=0.99 and *P*=0.995) would have required 530 mice.

$$1 - 0.995 = (0.99)^n$$

Therefore, infection of 24 mice with $5x10^6$ bacteria from the EXIT library resulted in a 99.5% probability of any individual clone in the library establishing infection in the spleen of at least one of the mice. We screened the EXIT library in two independent experiments, each consisting of 24 mice, which resulted in a 99.5% confidence for complete coverage of the library in each independent replicate. Unfortunately, we were unable to comprehensively evaluate the library in the lungs due to the unnecessarily high number of mice (530) that would have been required. Therefore, the analysis in Chapter 2 was primarily focused on the robust and comprehensive dataset from the spleens.

REFERENCES

Orme, I. & M. Gonzalez-Juarrero, (2007) Animal models of *M. tuberculosis* Infection. *Current protocols in microbiology* **Chapter 10**: Unit 10A 15.

Sambrook, J.E., E.F. Fritsch & T. Maniatis, (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.

Sassetti, C.M. & E.J. Rubin, (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**: 12989-12994.

Zhang, Y.J., M.C. Reddy, T.R. Ioerger, A.C. Rothchild, V. Dartois, B.M. Schuster, A. Trauner, D. Wallis, S. Galaviz, C. Huttenhower, J.C. Sacchettini, S.M. Behar & E.J. Rubin, (2013) Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. *Cell* **155**: 1296-1308.

APPENDIX II: EXIT RESULTS

Table 6.1 EXIT results

ORF number	Name	Product	Exported fusion junctions	ID in Lung	*in silico*	MS ID as exported	GR ID as exported	Predicted to be essential in vitro or during infection
Rv0011c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	83		TM	MEM 8, MEM 9, CW 14		
Rv0012		PROBABLE CONSERVED MEMBRANE PROTEIN	39, 46, 93, 102		TM	MEM 6, MEM 7, MEM 9, CW 14		Mouse 27
Rv0014c	pknB	TRANSMEMBRANE SERINE/THREONINE-PROTEIN KINASE B PKNB (PROTEIN KINASE B) (STPK B)	404, 421, 445, 446, 594		TM	CF 2, CF 3, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, SOL 16		in vitro 28, in vitro 29
Rv0015c	pknA	TRANSMEMBRANE SERINE/THREONINE-PROTEIN KINASE A PKNA (PROTEIN KINASE A) (STPK A)	362		TM	MEM 7, MEM 9, CW 14		in vitro 28, in vitro 29
Rv0016c	pbpA	PROBABLE PENICILLIN-BINDING PROTEIN PBPA	29, 51		SP, TM	MEM 8, MEM 9, CW 14		Mouse 27
Rv0017c	rodA, ftsW	PROBABLE CELL DIVISION PROTEIN RODA	412		TM	MEM 9, CW 14		Macrophage 23, Mouse 27
Rv0037c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	122, 349, 377		TM	MEM 8, MEM 9, CW 14		Macrophage 23
Rv0039c		POSSIBLE CONSERVED TRANSMEMBRANE PROTEIN	45, 50, 51, 58, 60, 95	Yes	SP, TM			
Rv0040c	mtc28	SECRETED PROLINE RICH PROTEIN MTC28 (PROLINE RICH 28 KDA ANTIGEN)	61, 85, 102, 178, 275, 283	Yes	SP, TM	CF 2, CF 3, MEM 7, CW 14		Mouse 27
Rv0048c		POSSIBLE MEMBRANE PROTEIN	123, 148, 201		TM	CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14		
Rv0062	celA1, celA	POSSIBLE CELLULASE CELA1 (ENDOGLUCANASE) (ENDO-1,4-BETA-GLUCANASE) (FI-CMCASE) (CARBOXYMETHYL CELLULASE)	91, 99, 106, 168, 173, 183, 191, 268, 308	Yes	TM	CF 2, MEM 7, CW 14		
Rv0064		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	275, 316, 423, 529	Yes	TM	MEM 7, CW 12	PhoA 18, BlaTEM 20	
Rv0072		PROBABLE GLUTAMINE-TRANSPORT TRANSMEMBRANE PROTEIN ABC TRANSPORTER	33, 59, 80, 83, 165, 275	Yes	TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14	BlaTEM 20	
Rv0076c		PROBABLE MEMBRANE PROTEIN	125		TM			
Rv0083		PROBABLE	303, 368,		TM	MEM 9		

		OXIDOREDUCTASE	385, 493, 498					
Rv0084	hycD, hevD	POSSIBLE FORMATE HYDROGENLYASE HYCD (FHL)	86, 174, 248, 250, 253, 306	Yes	TM			Macaque 22
Rv0086	hycQ	POSSIBLE HYDROGENASE HYCQ	58, 184, 189, 480	Yes	TM			Mouse 27, in vitro 28
Rv0092	ctpA	PROBABLE CATION TRANSPORTER P-TYPE ATPASE A CTPA	127, 374, 379, 444	Yes	TM	MEM 9, CW 12, CW 14	BlaTEM 20	Macrophage 23
Rv0093c		PROBABLE CONSERVED MEMBRANE PROTEIN	161, 189, 191, 214	Yes	TM	MEM 8, MEM 9, CW 12		
Rv0102		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	52, 71, 75, 136, 150, 156, 198, 213, 331, 416	Yes	TM	MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv0103c	ctpB	PROBABLE CATION-TRANSPORTER P-TYPE ATPASE B CTPB	130, 140, 213, 221	Yes	TM	MEM 9, CW 14		
Rv0110		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	72, 191		TM	MEM 8		
Rv0111		POSSIBLE TRANSMEMBRANE ACYLTRANSFERASE	70, 125, 428, 439, 457	Yes	TM	MEM 6, CW 11, CW 12		
Rv0116c		POSSIBLE CONSERVED MEMBRANE PROTEIN	18, 26, 82, 120, 192		SP, TM	CF 2, MEM 7	BlaTEM 20	
Rv0125	pepA, mtb32a	PROBABLE SERINE PROTEASE PEPA (SERINE PROTEINASE) (MTB32A)	45, 77, 109, 132, 134, 144, 186, 204, 266, 299, 304	Yes	SP, Tat SP, TM	CF 2, CF 3, MEM 7, MEM 8, MEM 9, CW 12, CW 13, CW 14, WCL 15, SOL 16	PhoA 18, BlaTEM 20	Mouse 27
Rv0128		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	39, 195		TM			
Rv0143c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	94, 97, 245, 334, 342	Yes	TM	MEM 8, MEM 9		
Rv0157	pntB	PROBABLE NAD(P) TRANSHYDROGENASE (SUBUNIT BETA) PNTB [INTEGRAL MEMBRANE PROTEIN] (PYRIDINE NUCLEOTIDE TRANSHYDROGENASE SUBUNIT BETA) (NICOTINAMIDE NUCLEOTIDE TRANSHYDROGENASE SUBUNIT BETA)	63, 190, 192, 253, 260	Yes	TM	MEM 4, MEM 5, MEM 9, MEM 10, CW 13, CW 14, WCL 15		Mouse 27
Rv0167	yrbE1A	CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBE1A	88, 90, 110, 112, 183, 194	Yes	TM	MEM 4, MEM 8, MEM 9, CW 14		
Rv0168	yrbE1B	CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBE1B	86, 108		TM	MEM 8, MEM 9, CW 14		
Rv0169	mce1A, mce1	MCE-FAMILY PROTEIN MCE1A	37, 50, 59, 116, 179, 400	Yes	TM	MEM 8, MEM 9, CW 12, CW 13, CW 14	BlaTEM 20	Macrophage 23, Macrophage 24, Mouse 25
Rv0170	mce1B, mceD	MCE-FAMILY PROTEIN MCE1B	39, 46, 154		TM	MEM 4, MEM 7, MEM 8, MEM 9, CW 13, CW 14, WCL 15	BlaTEM 20	Macrophage 23, Macrophage 24, Mouse 25
Rv0171	mce1C	MCE-FAMILY PROTEIN	48, 49, 60,	Yes	SP,	MEM 8, MEM 9,	BlaTEM	Macrophage

		MCE1C	146, 242, 338, 372, 449, 461, 489, 503, 504		TM	CW 13, CW 14	20	23, Mouse 25
Rv0172	mce1D	MCE-FAMILY PROTEIN MCE1D	472		TM	CF 2, CF 3, MEM 5, MEM 7, MEM 8, MEM 9, CW 12, CW 13, CW 14, WCL 15	BlaTEM 20	Macrophage 23
Rv0173	lprK, mce1E	POSSIBLE MCE-FAMILY LIPOPROTEIN LPRK (MCE-FAMILY LIPOPROTEIN MCE1E)	30, 33, 41, 42	Yes	Lipo, TM	MEM 5, MEM 7, MEM 8, MEM 9, CW 12, CW 13, CW 14, WCL 15	BlaTEM 20	Macaque 22, Macrophage 23, Mouse 25
Rv0174	mce1F	MCE-FAMILY PROTEIN MCE1F	46, 60, 161	Yes	TM	CF 2, CF 3, MEM 6, MEM 7, MEM 8, MEM 9, CW 12, CW 13, CW 14, WCL 15	BlaTEM 20	Macrophage 23
Rv0175		PROBABLE CONSERVED MCE ASSOCIATED MEMBRANE PROTEIN	66		TM	CF 3, MEM 5, MEM 9, MEM 10, CW 13, CW 14, WCL 15	PhoA 18, BlaTEM 20	Macrophage 23, Mouse 25
Rv0176		PROBABLE CONSERVED MCE ASSOCIATED TRANSMEMBRANE PROTEIN	279		TM	MEM 4, MEM 8, MEM 9, CW 14		Macrophage 23, Macrophage 24, Mouse 25
Rv0177		PROBABLE CONSERVED MCE ASSOCIATED PROTEIN	115		TM	MEM 4, MEM 8, MEM 9, CW 14		Macrophage 23, Macrophage 24, Mouse 27
Rv0178		PROBABLE CONSERVED MCE ASSOCIATED MEMBRANE PROTEIN	119, 162, 177	Yes	TM	CF 3, MEM 4, MEM 5, MEM 9, MEM 10, CW 13, CW 14, WCL 15	BlaTEM 20	Macrophage 23
Rv0179c	lprO	POSSIBLE LIPOPROTEIN LPRO	36, 40, 43, 67, 124, 159, 169, 183, 199, 273, 277, 296	Yes	SP, Lipo, TM	CW 14	PhoA 18	Macaque 22, Macrophage 23
Rv0180c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	40, 45, 66, 68, 79, 153, 180, 181, 191, 211	Yes	TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		Mouse 27
Rv0191		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	113, 310, 318, 382	Yes	TM	MEM 9		
Rv0192A		CONSERVED SECRETED PROTEIN	40, 46		SP, Tat SP, TM			
Rv0199		PROBABLE CONSERVED MEMBRANE PROTEIN	64, 70, 82, 155, 157	Yes	TM	MEM 5, MEM 9, MEM 10, CW 13, CW 14, WCL 15	BlaTEM 20	Mouse 25, Mouse 27
Rv0200		POSSIBLE CONSERVED TRANSMEMBRANE PROTEIN	22, 129, 145, 161, 172, 195		TM	MEM 4, MEM 9, CW 14		Macrophage 24
Rv0202c	mmpL 11	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL11	37, 39, 44, 166, 168, 407, 432, 433, 436	Yes	SP, TM	MEM 9, CW 12, CW 14		Macrophage 24
Rv0203		POSSIBLE EXPORTED PROTEIN	23, 34, 119	Yes	SP, Tat SP, TM	CF 2, CF 3, MEM 7		
Rv0205		PROBABLE CONSERVED TRANSMEMBRANE	87, 114, 127, 262,	Yes	TM	MEM 8, MEM 9		Mouse 27, in vitro 28

		PROTEIN	324, 348					
Rv0206c	mmpL 3	POSSIBLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL3	25, 40, 65, 76, 108, 117, 121, 163, 173, 205, 419, 444, 479	Yes	TM	CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15	PhoA 18	in vitro 29
Rv0218		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	64, 81, 90, 162, 179, 295		Tat SP, TM	MEM 8		Mouse 25
Rv0219		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	130		TM	MEM 7, MEM 9		
Rv0226c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	22, 169, 311, 343, 348	Yes	Tat SP, TM	MEM 8, MEM 9		in vitro 28, in vitro 29
Rv0227c		PROBABLE CONSERVED MEMBRANE PROTEIN	79, 219, 236, 326	Yes	SP, TM	CF 3, MEM 4, MEM 5, MEM 6, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		in vitro 28, in vitro 29
Rv0228		PROBABLE INTEGRAL MEMBRANE ACYLTRANSFERASE	212, 351		TM			in vitro 28, in vitro 29
Rv0236c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	24, 36, 191, 248, 329, 346, 425, 452, 510, 597, 1365	Yes	SP, TM	MEM 6, MEM 9		in vitro 28, in vitro 29
Rv0237	lpqI	PROBABLE CONSERVED LIPOPROTEIN LPQI	22, 26, 32, 36, 88, 111	Yes	SP, Lipo	CF 2, CF 3, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14		
Rv0249c		PROBABLE SUCCINATE DEHYDROGENASE [MEMBRANE ANCHOR SUBUNIT] (SUCCINIC DEHYDROGENASE)	80, 180	Yes	TM	MEM 8, MEM 9, MEM 10, CW 13, WCL 15		Mouse 25, Mouse 27, in vitro 29
Rv0261c	narK3	PROBABLE INTEGRAL MEMBRANE NITRITE EXTRUSION PROTEIN NARK3 (NITRITE FACILITATOR)	62, 331		TM			
Rv0283	eccB3	POSSIBLE CONSERVED MEMBRANE PROTEIN	111, 154, 422, 526	Yes	TM	CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		in vitro 28, in vitro 29
Rv0291	mycP3	PROBABLE MEMBRANE-ANCHORED MYCOSIN MYCP3 (SERINE PROTEASE) (SUBTILISIN-LIKE PROTEASE) (SUBTILASE-LIKE) (MYCOSIN-3)	433		SP, TM	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		in vitro 28, in vitro 29
Rv0309		POSSIBLE CONSERVED EXPORTED PROTEIN	48, 49, 61, 63, 70, 101, 105		SP, TM	CF 2, MEM 7, MEM 8, CW 12, CW 14		
Rv0312		CONSERVED HYPOTHETICAL PROLINE AND THREONINE RICH PROTEIN	479, 486, 551		TM	MEM 9, CW 14	BlaTEM 20	Mouse 27, in vitro 28
Rv0314c		POSSIBLE CONSERVED MEMBRANE PROTEIN	132, 139, 192, 211		TM	MEM 8, MEM 9, CW 14		
Rv0338c		PROBABLE IRON-SULFUR-	119, 203		TM	MEM 4, MEM 5,		in vitro 28, in

		BINDING REDUCTASE				MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		vitro 29
Rv0344c	lpqJ	PROBABLE LIPOPROTEIN LPQJ	50, 68, 71, 90, 126		SP, Lipo	MEM 9, CW 14		
Rv0346c	ansP2, aroP2	POSSIBLE L-ASPARAGINE PERMEASE ANSP2 (L-ASPARAGINE TRANSPORT PROTEIN)	52, 118, 234, 273, 312, 376		TM	MEM 8, MEM 9, CW 14	BlaTEM 20	Macrophage 24
Rv0359		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	42, 112		TM	MEM 8, MEM 9		
Rv0361		PROBABLE CONSERVED MEMBRANE PROTEIN	182, 186	Yes	TM	CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16	BlaTEM 20	Mouse 27
Rv0398c		POSSIBLE SECRETED PROTEIN	20, 24, 30, 41, 50, 206		SP, TM	CF 2, CF 3, MEM 7, SOL 16		
Rv0399c	lpqK	POSSIBLE CONSERVED LIPOPROTEIN LPQK	24		SP, Lipo	MEM 8, MEM 9, CW 14		in vitro 28
Rv0402c	mmpL 1	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL1	123, 788, 864		TM	MEM 7, MEM 8, MEM 9, CW 12, CW 14	BlaTEM 20	
Rv0403c	mmpS1	PROBABLE CONSERVED MEMBRANE PROTEIN MMPS1	40		SP, TM	CF 2, MEM 9		Mouse 27
Rv0411c	glnH	PROBABLE GLUTAMINE-BINDING LIPOPROTEIN GLNH (GLNBP)	25, 68	Yes	SP, Tat SP, Lipo	CF 2, MEM 7, MEM 8, MEM 9, MEM 10, CW 14		in vitro 28, in vitro 29
Rv0412c		POSSIBLE CONSERVED MEMBRANE PROTEIN	65, 90, 94, 103, 269		TM	MEM 5, MEM 8, MEM 9, CW 14	BlaTEM 20	Mouse 27, in vitro 28, in vitro 29
Rv0418	lpqL	PROBABLE LIPOPROTEIN AMINOPEPTIDASE LPQL	38, 105, 165, 188, 278, 459	Yes	SP, Lipo, TM	CF 3, MEM 4, MEM 5, MEM 8, MEM 9, CW 14	BlaTEM 20	
Rv0420c		POSSIBLE TRANSMEMBRANE PROTEIN	65, 92, 113, 131	Yes	Tat SP, TM			
Rv0426c		POSSIBLE TRANSMEMBRANE PROTEIN	83	Yes	SP, TM	MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		Macrophage 23
Rv0431	AT103	PUTATIVE TUBERCULIN RELATED PEPTIDE	71, 89, 100, 131, 155	Yes	TM	CF 2, CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14		in vitro 29
Rv0432	sodC	PROBABLE PERIPLASMIC SUPEROXIDE DISMUTASE [CU-ZN] SODC	50, 53, 66, 85, 102, 109, 130, 131, 134, 139, 140, 153, 164, 189, 224, 233	Yes	SP, Lipo	CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15	BlaTEM 20	Mouse 27
Rv0446c		POSSIBLE CONSERVED TRANSMEMBRANE PROTEIN	120, 123		TM	MEM 7, MEM 9		
Rv0450c	mmpL 4	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL4	149, 182, 185, 188, 239, 248, 301, 420, 427, 500		TM	CF 3, MEM 6, MEM 8, MEM 9, MEM 10, CW 12, CW 14	PhoA 18, BlaTEM 20	in vitro 28, in vitro 29
Rv0451c	mmpS4	PROBABLE CONSERVED MEMBRANE PROTEIN	25, 60, 138	Yes	TM	MEM 9		

		MMPS4						
Rv0455c		CONSERVED HYPOTHETICAL PROTEIN	19, 138		SP, TM	CF 2, CF 3, MEM 7, MEM 9, CW 13, CW 14, WCL 15, SOL 16	PhoA 18	in vitro 29
Rv0461		PROBABLE TRANSMEMBRANE PROTEIN	105, 149		TM	MEM 8, MEM 9		
Rv0476		POSSIBLE CONSERVED TRANSMEMBRANE PROTEIN	27, 67		SP, TM	MEM 8, MEM 9		
Rv0477		POSSIBLE CONSERVED SECRETED PROTEIN	28		SP,	CF 2, CF 3, MEM 7		
Rv0479c		PROBABLE CONSERVED MEMBRANE PROTEIN	125, 155, 180		TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		in vitro 28, in vitro 29
Rv0490	senX3	PUTATIVE TWO COMPONENT SENSOR HISTIDINE KINASE SENX3	121		TM	MEM 7, MEM 9, CW 14		Mouse 25
Rv0497		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	244		TM	MEM 8, MEM 9, CW 12, CW 14		Macrophage 24, Mouse 27, in vitro 29
Rv0506	mmpS2	PROBABLE CONSERVED MEMBRANE PROTEIN MMPS2	145		SP, TM	CF 2, MEM 7, MEM 8, MEM 9, CW 14	BlaTEM 20	
Rv0507	mmpL 2	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL2	188, 410		TM	CF 2, CW 12, CW 14		in vitro 29
Rv0517		POSSIBLE MEMBRANE ACYLTRANSFERASE	152, 248, 294, 299		TM	MEM 9, CW 14	BlaTEM 20	
Rv0518		POSSIBLE EXPORTED PROTEIN	48, 53, 61, 85, 90, 101	Yes		CF 2, CF 3, MEM 8		
Rv0522	gabP	PROBABLE GABA PERMEASE GABP (4-AMINO BUTYRATE TRANSPORT CARRIER) (GAMA-AMINOBUTYRATE PERMEASE)	149, 251, 271, 400		TM			
Rv0526		POSSIBLE THIOREDOXIN PROTEIN (THIOL-DISULFIDE INTERCHANGE PROTEIN)	31, 39, 47, 48	Yes	SP, Tat SP, Lipo	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv0528		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	34, 42, 190		SP, TM	MEM 9, CW 14		in vitro 28, in vitro 29
Rv0529	ccsA, ccsB	POSSIBLE CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCSA	39, 41, 50, 78, 151, 159, 188	Yes	TM	MEM 9, CW 14		in vitro 28, in vitro 29
Rv0534c	menA	1,4-DIHYDROXY-2-NAPHTHOATE OCTAPRENYLTRANSFERASE MENA (DHNA-OCTAPRENYLTRANSFERASE)	65, 102, 123, 243, 251, 256	Yes	TM	MEM 9, CW 14		in vitro 29
Rv0537c		PROBABLE INTEGRAL MEMBRANE PROTEIN	65, 98, 104, 116, 158, 177, 266	Yes	Tat SP, TM	MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		
Rv0541c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	45, 47, 111		TM			Mouse 27, in vitro 28, in vitro 29
Rv0545c	pitA	PROBABLE LOW-AFFINITY INORGANIC PHOSPHATE TRANSPORTER INTEGRAL MEMBRANE PROTEIN	26, 36, 128, 161, 170, 174	Yes	TM	MEM 9		Mouse 27

		PITA						
Rv0559c		POSSIBLE CONSERVED SECRETED PROTEIN	79		SP,	CF 2, CF 3, MEM 7, MEM 8, CW 12, CW 13		
Rv0583c	lpqN	PROBABLE CONSERVED LIPOPROTEIN LPQN	79, 104, 111, 128, 133, 141, 148, 215	Yes	SP, Lipo	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16	BlaTEM 20	
Rv0585c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	622, 642, 713		TM	MEM 8		
Rv0587	yrbE2A	CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBE2A	88, 90		TM			
Rv0588	yrbE2B	CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBE2B	114		TM			Mouse 27, in vitro 28
Rv0590	mce2B	MCE-FAMILY PROTEIN MCE2B	36, 39, 50	Yes	TM	MEM 9		Mouse 27
Rv0593	lprL, mce2E	POSSIBLE MCE-FAMILY LIPOPROTEIN LPRL (MCE-FAMILY LIPOPROTEIN MCE2E)	261		Lipo, TM			
Rv0594	mce2F	MCE-FAMILY PROTEIN MCE2F	48, 141		TM	MEM 9	BlaTEM 20	
Rv0601c		PROBABLE TWO COMPONENT SENSOR KINASE [FIRST PART]	18, 29, 36	Yes	TM			
Rv0603		POSSIBLE EXPORTED PROTEIN	20, 35, 44, 54, 55, 62, 66, 101		SP, TM	CF 3		
Rv0604	lpqO	PROBABLE CONSERVED LIPOPROTEIN LPQO	20, 34, 59, 88, 128, 129, 134, 153, 177, 182, 191, 201, 206, 267	Yes	SP, Lipo	MEM 4, MEM 8, MEM 9, CW 14		
Rv0615		PROBABLE INTEGRAL MEMBRANE PROTEIN	47, 63	Yes	TM			
Rv0621		POSSIBLE MEMBRANE PROTEIN	108, 112, 260	Yes	TM	MEM 9		
Rv0622		POSSIBLE MEMBRANE PROTEIN	144, 164		TM	CW 12		Mouse 27
Rv0625c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	50	Yes	TM	MEM 8, CW 14		
Rv0658c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	38, 45, 53, 59, 60, 109, 118	Yes	TM			
Rv0671	lpqP	POSSIBLE CONSERVED LIPOPROTEIN LPQP	41, 65, 104, 111, 203, 258		SP, Lipo, TM	CW 14		Macaque 22, Mouse 27
Rv0676c	mmpL5	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL5	179, 243, 272, 421, 432, 758		TM	MEM 5, MEM 8, MEM 9, CW 14	PhoA 18, BlaTEM 20	
Rv0677c	mmpS5	POSSIBLE CONSERVED MEMBRANE PROTEIN MMPS5	20, 60, 62, 140	Yes	TM	CF 2, CF 3, MEM 7, MEM 8, MEM 9, CW 14	BlaTEM 20	Macrophage 23
Rv0679c		CONSERVED HYPOTHETICAL THREONINE RICH PROTEIN	45, 47, 58, 65, 70, 87	Yes	SP, Lipo	CF 3, MEM 8, MEM 9, CW 14		

Rv0680c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	21, 37		SP, TM	CF 2, MEM 7, MEM 8, CW 14		
Rv0713		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	136, 178		TM	MEM 9, CW 14		
Rv0732	secY	PROBABLE PREPROTEIN TRANSLOCASE SECY	31, 36, 55, 66, 67, 85, 140, 212, 295, 398	Yes	TM	MEM 7, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		in vitro 28, in vitro 29
Rv0774c		PROBABLE CONSERVED EXPORTED PROTEIN	40		SP, Tat SP, TM	CF 2, MEM 7	BlaC 19	
Rv0779c		POSSIBLE CONSERVED TRANSMEMBRANE PROTEIN	48, 60, 65, 85, 95, 97, 105	Yes	TM	MEM 9, CW 14		
Rv0783c	emrB	POSSIBLE MULTIDRUG RESISTANCE INTEGRAL MEMBRANE EFFLUX PROTEIN EMRB	352, 395, 520, 538	Yes	TM	MEM 9	BlaTEM 20	
Rv0787		HYPOTHETICAL PROTEIN	136			CF 2, CF 3, MEM 7, MEM 9		
Rv0817c		PROBABLE CONSERVED EXPORTED PROTEIN	83, 140, 151		SP, TM	MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv0822c		CONSERVED HYPOTHETICAL PROTEIN	188, 218, 235, 243, 250, 471, 565	Yes		CF 2, CW 14		
Rv0835	lpqQ	POSSIBLE LIPOPROTEIN LPQQ	33, 60		SP,	CF 2, MEM 7, MEM 9		Mouse 27
Rv0838	lpqR	PROBABLE CONSERVED LIPOPROTEIN LPQR	50, 63, 69, 75, 88, 95, 241, 243	Yes	SP, Lipo	CF 2, MEM 7, MEM 8		
Rv0846c		PROBABLE OXIDASE	41, 43, 44, 47		SP, Tat SP, Lipo	MEM 8, MEM 9, CW 14	BlaC 19	
Rv0870c		POSSIBLE CONSERVED INTEGRAL MEMBRANE PROTEIN	60, 61, 63, 88		TM	MEM 8, MEM 9, CW 14		
Rv0875c		POSSIBLE CONSERVED EXPORTED PROTEIN	19, 21, 28, 101		SP, TM	MEM 4, MEM 9, CW 14		Mouse 27, in vitro 28, in vitro 29
Rv0879c		POSSIBLE CONSERVED TRANSMEMBRANE PROTEIN	53, 70		TM	MEM 8, MEM 9		
Rv0888		PROBABLE EXPORTED PROTEIN	30, 53, 75, 143, 146		SP, TM	CF 2, MEM 8, MEM 9, CW 14		Macrophage 24
Rv0892		PROBABLE MONOOXYGENASE	41		TM	MEM 9, CW 12, CW 14		
Rv0899	ompA	OUTER MEMBRANE PROTEIN A OMPA	56, 59, 86, 100, 142, 287	Yes	TM	CF 2, CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 14		
Rv0902c	prrB	TWO COMPONENT SENSOR HISTIDINE KINASE PRRB	110		SP, TM	MEM 5, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		in vitro 29
Rv0907		CONSERVED HYPOTHETICAL PROTEIN	24, 35, 51, 73, 76, 85	Yes		CF 2, CF 3, MEM 5, MEM 8, MEM 9, CW 14		Mouse 27, in vitro 29
Rv0912		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	31, 118, 119, 120, 135	Yes	TM			
Rv0917	betP	POSSIBLE GLYCINE BETAINE TRANSPORT INTEGRAL MEMBRANE PROTEIN BETP	53, 228, 315		TM		BlaTEM 20	

Rv0924c	mntH, Nramp, Mramp	DIVALENT CATION-TRANSPORT INTEGRAL MEMBRANE PROTEIN MNTH (BRAMP) (MRAMP)	183, 275, 426		TM	MEM 9		Macaque 22
Rv0928	pstS3, phoS2	PERIPLASMIC PHOSPHATE-BINDING LIPOPROTEIN PSTS3 (PBP-3) (PSTS3) (PHOS1)	20, 29, 44, 74, 118, 165, 202, 280, 285, 304	Yes	SP, Lipo	CF 2, CF 3, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		Macrophage 23, Mouse 27
Rv0931c	pknD, mbk	TRANSMEMBRANE SERINE/THREONINE-PROTEIN KINASE D PKND (PROTEIN KINASE D) (STPK D)	411, 419, 593, 653		TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16	BlaTEM 20	
Rv0932c	pstS2	PERIPLASMIC PHOSPHATE-BINDING LIPOPROTEIN PSTS2 (PBP-2) (PSTS2)	22, 34, 98, 151, 172, 190, 280, 326	Yes	SP, Lipo	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15, SOL 16		
Rv0934	pstS1, phoS1, phoS	PERIPLASMIC PHOSPHATE-BINDING LIPOPROTEIN PSTS1 (PBP-1) (PSTS1)	29, 36, 38, 45, 55, 170, 188, 210, 235	Yes	SP, Lipo	CF 1, CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16	PhoA 18, BlaTEM 20	
Rv0935	pstC1	PHOSPHATE-TRANSPORT INTEGRAL MEMBRANE ABC TRANSPORTER PSTC1	172, 184		TM	MEM 8, MEM 9, MEM 10		
Rv0936	pstA2	PHOSPHATE-TRANSPORT INTEGRAL MEMBRANE ABC TRANSPORTER PSTA2	239	Yes	TM	MEM 8, MEM 9, CW 14		
Rv0950c		CONSERVED HYPOTHETICAL PROTEIN	143, 222	Yes				Mouse 25, Mouse 27
Rv0954		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	73, 90, 129, 159		TM	MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, WCL 15		Macaque 22
Rv0961		PROBABLE INTEGRAL MEMBRANE PROTEIN	31		TM			
Rv0962c	lprP	POSSIBLE LIPOPROTEIN LPRP	33		SP, Lipo			
Rv0969	ctpV	PROBABLE METAL CATION TRANSPORTER P-TYPE ATPASE CTPV	169		TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		
Rv0982	mprB	PROBABLE TWO COMPONENT SENSOR KINASE MPRB	48, 82, 94, 100, 186	Yes	SP, TM	MEM 7, MEM 9, CW 14		in vitro 28, in vitro 29
Rv0983	pepD, mtb32b	PROBABLE SERINE PROTEASE PEPD (SERINE PROTEINASE) (MTB32B)	116, 119, 130, 146, 245, 292, 313, 316, 370, 398, 402	Yes	TM	CF 2, CF 3, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15, SOL 16	BlaTEM 20	Mouse 27
Rv0987		PROBABLE ADHESION COMPONENT TRANSPORT TRANSMEMBRANE PROTEIN ABC TRANSPORTER	58, 62, 73, 92, 533, 537		TM	MEM 6, MEM 9, CW 14		in vitro 29
Rv0988		POSSIBLE CONSERVED EXPORTED PROTEIN	45	Yes	SP, TM	CF 2, MEM 8, MEM 9, CW 14		Mouse 27, in vitro 29
Rv0999		HYPOTHETICAL PROTEIN	57, 60, 89, 93, 169, 201, 208	Yes	SP,	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, CW 14		

Rv1002c		CONSERVED MEMBRANE PROTEIN	75, 149, 285, 328, 483		TM	MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv1004c		PROBABLE MEMBRANE PROTEIN	44, 49, 64, 71, 113, 119, 122, 168, 172, 192, 211, 218, 230, 237, 245, 287, 408	Yes	SP, TM	MEM 8	BlaTEM 20	
Rv1006		HYPOTHETICAL PROTEIN	33, 56, 72, 81, 128, 219, 263, 286, 305, 361, 400, 504	Yes	SP,	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15	PhoA 18	
Rv1009	rpfB	Probable resuscitation-promoting factor rpfB	15, 21, 33, 63, 65, 98, 117, 128, 145, 165, 177, 247, 271, 288, 292, 300	Yes	SP, Lipo, TM		BlaTEM 20	in vitro growth-defect 28
Rv1016c	lpqT	PROBABLE CONSERVED LIPOPROTEIN LPQT	69		SP, Lipo	CF 2, CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 14		Macrophage 23, Mouse 25
Rv1022	lpqU	PROBABLE CONSERVED LIPOPROTEIN LPQU	39, 51, 54	Yes	SP, TM	MEM 8, MEM 9, CW 14		
Rv1024		POSSIBLE CONSERVED MEMBRANE PROTEIN	187, 209, 225	Yes	Tat SP	MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv1026		CONSERVED HYPOTHETICAL PROTEIN	245					in vitro 28
Rv1029	kdpA	Probable Potassium-transporting ATPase A chain KDPA (Potassium-translocating ATPase A chain) (ATP phosphohydrolase [potassium-transporting] A chain) (Potassium binding and translocating subunit A)	37, 220, 236, 244, 327, 369, 443, 451, 483	Yes	TM	CW 12		
Rv1030	kdpB	Probable Potassium-transporting P-type ATPase B chain KDPB (Potassium-translocating ATPase B chain) (ATP phosphohydrolase [potassium-transporting] B chain) (Potassium binding and translocating subunit B)	88, 262, 263		TM	MEM 8, MEM 9, CW 14		
Rv1031	kdpC	Probable Potassium-transporting ATPase C chain KDPC (Potassium-translocating ATPase C chain) (ATP phosphohydrolase [potassium-transporting] C chain) (Potassium binding and translocating subunit C)	27, 52, 60		TM	MEM 8, MEM 9, CW 14		
Rv1032c	trcS	TWO COMPONENT SENSOR HISTIDINE KINASE TRCS	51, 82		TM			
Rv1064c	lpqV	POSSIBLE LIPOPROTEIN LPQV	23, 27, 33, 40, 53, 62		SP, Lipo			
Rv1072		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	137, 200, 210		TM	MEM 8, MEM 9, CW 14	PhoA 18	in vitro 29
Rv1081c		PROBABLE CONSERVED MEMBRANE PROTEIN	54	Yes	TM	MEM 9		in vitro 29
Rv1085c		POSSIBLE HEMOLYSIN-	63, 181	Yes	TM	MEM 9		

		LIKE PROTEIN						
Rv1096		POSSIBLE GLYCOSYL HYDROLASE	31, 84, 87, 119, 133, 143, 144, 154, 284		TM	CF 2, CF 3, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14	BlaTEM 20	Macrophage 23, Mouse 27, in vitro growth-defect 28
Rv1097c		PROBABLE MEMBRANE GLYCINE AND PROLINE RICH PROTEIN	114, 207		TM	CF 2, CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14		Mouse 27
Rv1100		CONSERVED HYPOTHETICAL PROTEIN	67, 101, 107, 110, 122, 230	Yes	TM	CF 3, MEM 8, MEM 9, CW 12, CW 14		Mouse 27, in vitro growth-defect 28
Rv1111c		CONSERVED HYPOTHETICAL PROTEIN	34, 44, 56		TM	MEM 9, CW 14		Mouse 25, Mouse 27, in vitro 29
Rv1132		CONSERVED MEMBRANE PROTEIN	166, 180, 318		TM	MEM 8, MEM 9, CW 14		
Rv1140		PROBABLE INTEGRAL MEMBRANE PROTEIN	101, 135	Yes	TM	MEM 8, MEM 9, CW 14		
Rv1145	mmpL 13a	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL13A	39, 63		TM	MEM 9		
Rv1158c		CONSERVED HYPOTHETICAL ALA-, PRO-RICH PROTEIN	23		SP, TM	MEM 7		
Rv1159	pimE	CONSERVED TRANSMEMBRANE PROTEIN	277, 288		TM			Mouse 27, in vitro 29
Rv1166	lpqW	PROBABLE CONSERVED LIPOPROTEIN LPQW	25, 63, 78, 99, 112, 160, 161, 174, 187, 211, 251, 297, 485	Yes	SP, Lipo	CF 2, MEM 7, MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv1174c	TB8.4	LOW MOLECULAR WEIGHT T-CELL ANTIGEN TB8.4	26, 55, 83	Yes	SP,	CF 2, CF 3, MEM 7, MEM 10, CW 13, WCL 15, SOL 16	BlaTEM 20	
Rv1183	mmpL 10	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL10	50, 329, 862, 932		SP, TM	MEM 6, MEM 7, MEM 8, MEM 9, CW 14		Mouse 25
Rv1184c		POSSIBLE EXPORTED PROTEIN	16, 21, 28, 48, 88, 89, 134, 174, 190, 218, 253, 255, 354, 357		SP, TM	MEM 8, CW 12, CW 14		Mouse 25
Rv1200		PROBABLE CONSERVED INTEGRAL MEMBRANE TRANSPORT PROTEIN	42, 102, 269, 323		TM			Macaque 22
Rv1217c		PROBABLE TETRONASIN-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER	67, 70, 102, 174, 227, 352, 448, 513, 518	Yes	TM	MEM 8, MEM 9, CW 14		
Rv1223	htrA, degP	PROBABLE SERINE PROTEASE HTRA (DEGP PROTEIN)	191, 194, 225, 254, 261, 294, 386, 387, 402, 403, 466, 470	Yes	TM	CF 2, CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		in vitro 28, in vitro 29
Rv1226c		PROBABLE TRANSMEMBRANE PROTEIN	197, 206, 210	Yes	TM	MEM 8, MEM 9, CW 14		
Rv1228	lpqX	PROBABLE LIPOPROTEIN	37, 56, 68,	Yes	Lipo	CF 2, MEM 9,		

		LPQX	74			CW 14		
Rv1230c		POSSIBLE MEMBRANE PROTEIN	53, 69, 74, 118, 130, 270, 271, 275, 305	Yes	SP, TM	CW 12	BlaTEM 20	
Rv1234		PROBABLE TRANSMEMBRANE PROTEIN	112, 116		TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		Macaque 22, Mouse 27
Rv1236	sugA	PROBABLE SUGAR-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER SUGA	146, 246		Tat SP, TM	MEM 8, MEM 9		Macrophage 23, Macrophage 24, Mouse 25, Mouse 27
Rv1237	sugB	PROBABLE SUGAR-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER SUGB	212, 219, 228, 237		TM	MEM 8, MEM 9, CW 14		Macrophage 23, Mouse 25
Rv1244	lpqZ	PROBABLE LIPOPROTEIN LPQZ	26, 47, 65, 85, 123, 137, 154, 249, 274, 275		SP, Lipo	MEM 8, MEM 9, CW 12, CW 14		Macrophage 23, Mouse 25
Rv1250		PROBABLE DRUG-TRANSPORT INTEGRAL MEMBRANE PROTEIN	104, 106, 111, 120, 386, 396, 438, 440, 441		TM			
Rv1252c	lprE	PROBABLE LIPOPROTEIN LPRE	61, 72, 130, 157, 199, 200	Yes	SP, Lipo	CF 2, MEM 7, MEM 8, MEM 9, CW 14		
Rv1258c		PROBABLE CONSERVED INTEGRAL MEMBRANE TRANSPORT PROTEIN	372		TM			
Rv1266c	pknH	PROBABLE TRANSMEMBRANE SERINE/THREONINE-PROTEIN KINASE H PKNH (PROTEIN KINASE H) (STPK H)	603	Yes	TM	MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		
Rv1269c		CONSERVED PROBABLE SECRETED PROTEIN	29, 34, 44	Yes	SP, Tat SP	CF 2, CF 3, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15	PhoA 17	
Rv1270c	lprA	POSSIBLE LIPOPROTEIN LPRA	22, 30, 36, 60, 93, 123, 161, 238	Yes	SP, Lipo, TM	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		
Rv1271c		CONSERVED HYPOTHETICAL SECRETED PROTEIN	28, 30		SP, TM	CF 3		
Rv1273c		PROBABLE DRUGS-TRANSPORT TRANSMEMBRANE ATP-BINDING PROTEIN ABC TRANSPORTER	38, 50, 64		SP, TM	MEM 8, MEM 9, CW 12		Macrophage 23, Mouse 27
Rv1274	lprB	POSSIBLE LIPOPROTEIN LPRB	27, 55, 69, 74, 86, 127	Yes	SP, Lipo	MEM 8, MEM 9, MEM 10, CW 13, WCL 15		in vitro 28
Rv1275	lprC	POSSIBLE LIPOPROTEIN LPRC	18, 27	Yes	SP, Lipo	CF 2, MEM 4, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		
Rv1280c	oppA	PROBABLE PERIPLASMIC	49, 54, 81,	Yes	Tat	MEM 5, MEM 8,		

		OLIGOPEPTIDE-BINDING LIPOPROTEIN OPPA	252, 267, 400, 469		SP, Lipo, TM	MEM 9, MEM 10, CW 14		
Rv1283c	oppB	PROBABLE OLIGOPEPTIDE-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER OPPB	82, 111, 169, 174, 177, 178, 258, 262, 282	Yes	TM	MEM 9, CW 14		
Rv1290c		CONSERVED HYPOTHETICAL PROTEIN	94, 184, 194	Yes	Tat SP, TM	MEM 9, CW 14		
Rv1302	rfe, wecA	PROBABLE UNDECAPAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE RFE (UDP-GlcNAc TRANSFERASE)	106, 109, 114, 237, 360, 371, 382	Yes	TM	MEM 8, MEM 9		in vitro 29
Rv1304	atpB	PROBABLE ATP SYNTHASE A CHAIN ATPB (PROTEIN 6)	145		TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, WCL 15		Mouse 25, in vitro 29
Rv1320c		POSSIBLE ADENYLATE CYCLASE (ATP PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE)	151, 237	Yes	TM	MEM 9, CW 14		
Rv1348		PROBABLE DRUGS-TRANSPORT TRANSMEMBRANE ATP-BINDING PROTEIN ABC TRANSPORTER	322, 342	Yes	TM	MEM 4, MEM 9		Mouse 27, in vitro 28, in vitro 29
Rv1352		CONSERVED HYPOTHETICAL PROTEIN	31, 36		SP, TM	CF 2, CF 3, MEM 7		
Rv1362c		POSSIBLE MEMBRANE PROTEIN	95, 97, 103, 126, 132, 159		TM	MEM 4, MEM 9, CW 14		
Rv1363c		POSSIBLE MEMBRANE PROTEIN	119, 122, 139, 144, 173, 214, 228	Yes	TM	MEM 8, MEM 9, CW 14		
Rv1368	lprF	PROBABLE CONSERVED LIPOPROTEIN LPRF	38, 50, 136, 226	Yes	SP, Lipo, TM	CF 2, CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14	BlaTEM 20	
Rv1411c	lprG, P27	PROBABLE CONSERVED LIPOPROTEIN LPRG	24, 33, 168, 198		SP, Lipo	CF 2, CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		Macrophage 23, Macrophage 24, Mouse 25, Mouse 27
Rv1418	lprH	PROBABLE LIPOPROTEIN LPRH	43, 48, 208	Yes	SP, Lipo, TM	MEM 8, MEM 9, CW 14		
Rv1419		HYPOTHETICAL PROTEIN	85, 130		SP,	CF 2, CF 3, MEM 7	BlaTEM 20	
Rv1424c		POSSIBLE MEMBRANE PROTEIN	65, 67, 191, 217, 223	Yes	SP,	MEM 8, MEM 9, CW 12, CW 14		
Rv1431		CONSERVED MEMBRANE PROTEIN	65, 164, 175, 177	Yes	TM	MEM 8, MEM 9, CW 14		
Rv1433		POSSIBLE CONSERVED EXPORTED PROTEIN	44, 84, 124		SP, TM			Macaque 22
Rv1435c		Probable conserved Proline, Glycine, Valine-rich secreted protein	71, 117, 167, 180	Yes	SP,	CF 2, MEM 7	BlaTEM 20	
Rv1451	ctaB	PROBABLE CYTOCHROME C OXIDASE ASSEMBLY FACTOR CTAB	47, 61, 65, 67, 87, 271	Yes	TM	MEM 9, CW 12		in vitro 29
Rv1456c		PROBABLE UNIDENTIFIED	30, 37,	Yes	TM	MEM 9		in vitro 28, in

		ANTIBIOTIC-TRANSPORT INTEGRAL MEMBRANE ABC TRANSPORTER	261, 270, 282, 291					vitro 29
Rv1457c		PROBABLE UNIDENTIFIED ANTIBIOTIC-TRANSPORT INTEGRAL MEMBRANE ABC TRANSPORTER	71, 163		TM			in vitro 28, in vitro 29
Rv1459c		POSSIBLE CONSERVED INTEGRAL MEMBRANE PROTEIN	158, 190, 388, 400, 418, 520		TM	MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv1477		HYPOTHETICAL INVASION PROTEIN	46, 279, 360, 373	Yes	SP,	CF 2, CF 3, MEM 7, MEM 9, CW 13, CW 14		in vitro 28, in vitro 29
Rv1478		HYPOTHETICAL INVASION PROTEIN	27, 32, 48	Yes	SP,			
Rv1487		CONSERVED MEMBRANE PROTEIN	20, 29, 34	Yes	TM	MEM 8, MEM 9, CW 14		
Rv1490		PROBABLE MEMBRANE PROTEIN	160, 342, 426		TM			Mouse 27, in vitro 28, in vitro 29
Rv1491c		CONSERVED MEMBRANE PROTEIN	92, 110, 123, 202, 204	Yes	TM			
Rv1508c		Probable membrane protein	232		TM	MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		
Rv1510		conserved probable membrane protein	205, 272, 349		TM			
Rv1517		CONSERVED HYPOTHETICAL TRANSMEMBRANE PROTEIN	69		Tat SP, TM			
Rv1522c	mmpL 12	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL12	48, 449, 517, 953	Yes	SP, TM	MEM 9, CW 12		in vitro 29
Rv1539	lspA	PROBABLE LIPOPROTEIN SIGNAL PEPTIDASE LSPA	59, 86	Yes	TM	MEM 9		in vitro 28, in vitro 29
Rv1541c	lprI	Possible lipoprotein lprI	27, 28, 60	Yes	SP, Lipo	MEM 7		
Rv1554	frdC	PROBABLE FUMARATE REDUCTASE [MEMBRANE ANCHOR SUBUNIT] FRDC (FUMARATE DEHYDROGENASE) (FUMARIC HYDROGENASE)	48		TM			
Rv1565c		CONSERVED HYPOTHETICAL MEMBRANE PROTEIN	71, 151, 226, 263, 440		TM	MEM 9		Mouse 27, in vitro 29
Rv1566c		Possible inv protein	22, 24, 48, 60, 69, 71, 83, 93		SP, TM	MEM 8, CW 14	PhoA 17, PhoA 18	Mouse 27
Rv1607	chaA	Probable ionic transporter integral membrane protein chaA	83, 84, 166, 168, 244	Yes	TM	MEM 9		
Rv1610		POSSIBLE CONSERVED MEMBRANE PROTEIN	57, 117		TM	MEM 8, MEM 9, CW 14		in vitro 28
Rv1616		CONSERVED MEMBRANE PROTEIN	51		TM			
Rv1619		CONSERVED MEMBRANE PROTEIN	65, 134		TM	MEM 9		
Rv1622c	cydB	Probable integral membrane cytochrome D ubiquinol oxidase (subunit II) cydB (Cytochrome BD-I oxidase subunit II)	172		TM	MEM 8, MEM 9		Mouse 27, in vitro 28, in vitro 29
Rv1623c	cydA, appC	Probable integral membrane cytochrome D ubiquinol	44, 67, 138, 161,	Yes	TM	MEM 8, MEM 9, CW 12, CW 14		Mouse 27, in vitro 29

		oxidase (subunit I) cydA (Cytochrome BD-I oxidase subunit I)	177, 242, 389, 408, 425					
Rv1625c	cya	MEMBRANE-ANCHORED ADENYLYL CYCLASE CYA (ATP PYROPHOSPHATE-LYASE) (ADENYLATE CYCLASE)	70, 82, 132, 188		TM	MEM 9, CW 14		
Rv1635c		Probable conserved transmembrane protein	242, 397, 432, 446		TM		BlaTEM 20	
Rv1639c		CONSERVED HYPOTHETICAL MEMBRANE PROTEIN	77, 146, 153, 227, 239, 265, 338, 379, 422, 484	Yes	TM			
Rv1640c	lysX	Possible lysyl-tRNA synthetase 2 lysX	162, 168, 171, 262		TM	MEM 6, MEM 9, CW 12, CW 14		Macrophage 23, Mouse 25
Rv1648		Probable transmembrane protein	113		SP, TM	CW 14	BlaTEM 20	
Rv1672c		PROBABLE CONSERVED INTEGRAL MEMBRANE TRANSPORT PROTEIN	53, 57, 59, 72, 125, 172, 175, 271, 275, 386, 393, 404	Yes	TM	MEM 8		
Rv1677	dsbF	PROBABLE CONSERVED LIPOPROTEIN DSBF	32, 36, 48, 96		SP, Lipo	CF 3, MEM 9, MEM 10		
Rv1678		PROBABLE INTEGRAL MEMBRANE PROTEIN	202, 220, 235, 256, 270, 275		TM	MEM 9, CW 14		
Rv1686c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER	23, 27, 31	Yes	TM	MEM 8, MEM 9		
Rv1704c	cycA	PROBABLE D-SERINE/ALANINE/GLYCIN E TRANSPORTER PROTEIN CYCA	120, 436, 438		TM	MEM 9		Macaque 22
Rv1707		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	28, 30, 46		TM	MEM 8, MEM 9, CW 14	BlaTEM 20	
Rv1728c		CONSERVED HYPOTHETICAL PROTEIN	183, 234, 239	Yes			BlaTEM 20	Macrophage 24
Rv1733c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	59, 62, 64, 83, 85, 100, 146		SP, TM			
Rv1736c	narX	PROBABLE NITRATE REDUCTASE NARX	492, 587		TM	MEM 9, CW 14		
Rv1737c	narK2	POSSIBLE NITRATE/NITRITE TRANSPORTER NARK2	247, 248		TM	MEM 8, MEM 9		Mouse 27
Rv1739c		PROBABLE SULPHATE-TRANSPORT TRANSMEMBRANE PROTEIN ABC TRANSPORTER	158, 238, 261		TM			
Rv1743	pknE	PROBABLE TRANSMEMBRANE SERINE/THREONINE-PROTEIN KINASE E PKNE (PROTEIN KINASE E) (STPK E)			TM	CF 3, MEM 9, MEM 10, CW 13, CW 14, WCL 15, SOL 16	BlaTEM 20	
Rv1747		PROBABLE CONSERVED TRANSMEMBRANE ATP-BINDING PROTEIN ABC TRANSPORTER	722, 781, 820		TM	MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		
Rv1749c		POSSIBLE INTEGRAL MEMBRANE PROTEIN	86, 89, 113	Yes	TM	MEM 8, MEM 9, MEM 10, CW 13, WCL 15		

Rv1754c		CONSERVED HYPOTHETICAL PROTEIN	95, 98, 129, 301, 394	Yes	TM	MEM 8, MEM 9, CW 14		
Rv1779c		HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN	128, 174, 200, 278	Yes	TM	MEM 8, MEM 9	BlaTEM 20	
Rv1782	eccB5	PROBABLE CONSERVED MEMBRANE PROTEIN	92, 116, 135, 154, 380	Yes	TM	CF 2, MEM 4, MEM 8, MEM 9, CW 12, CW 14		in vitro 29
Rv1804c		CONSERVED HYPOTHETICAL PROTEIN	24		SP,	MEM 7		Macaque 22
Rv1810		CONSERVED HYPOTHETICAL PROTEIN	31, 41, 45, 50, 63, 89	Yes	SP, TM	CF 2, CF 3, MEM 7		Macaque 22
Rv1811	mgtC	POSSIBLE Mg2+ TRANSPORT P-TYPE ATPASE C MGTC	75, 117		TM	MEM 9		
Rv1813c		CONSERVED HYPOTHETICAL PROTEIN	47		SP, Tat SP, TM	CF 3, MEM 9, CW 14		
Rv1814	erg3	MEMBRANE-BOUND C-5 STEROL DESATURASE ERG3 (STEROL-C5-DESATURASE)	86		TM			
Rv1823		CONSERVED HYPOTHETICAL PROTEIN	79, 94, 98, 112, 188, 264, 292			MEM 8, MEM 9, CW 14		Mouse 27
Rv1824		CONSERVED HYPOTHETICAL MEMBRANE PROTEIN	29, 45, 96		TM	MEM 9		in vitro growth-defect 28
Rv1825		CONSERVED HYPOTHETICAL PROTEIN	92, 135, 163, 176, 276	Yes	TM	CF 2, CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 14	PhoA 18	Mouse 27
Rv1832	gcvB	Probable glycine dehydrogenase gcvB (Glycine decarboxylase) (Glycine cleavage system P-protein)	257			CF 2, MEM 4, MEM 8, MEM 9, WCL 15, SOL 16		in vitro 28, in vitro 29
Rv1842c		CONSERVED HYPOTHETICAL MEMBRANE PROTEIN	30, 42		TM	MEM 9, CW 14		
Rv1845c		CONSERVED HYPOTHETICAL TRANSMEMBRANE PROTEIN	51, 62, 76	Yes	SP, TM	MEM 7, MEM 9		Mouse 27, in vitro growth-defect 28, in vitro 29
Rv1857	modA	PROBABLE MOLYBDATE-BINDING LIPOPROTEIN MODA	26, 31, 87, 96, 109, 124, 187	Yes	SP, Lipo	MEM 8		
Rv1860	apa, mpt32, modD	ALANINE AND PROLINE RICH SECRETED PROTEIN APA (FIBRONECTIN ATTACHMENT PROTEIN) (Immunogenic protein MPT32) (Antigen MPT-32) (45-kDa glycoprotein) (45/47 kDa antigen)	35, 50, 73, 146	Yes	SP, TM	CF 1, CF 2, CF 3, MEM 7, MEM 8, CW 13, CW 14, WCL 15, SOL 16	BlaTEM 20	Mouse 27
Rv1861		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	74		TM	MEM 9		
Rv1863c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	59, 125, 148	Yes	TM			
Rv1877		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	128, 134, 234		TM			
Rv1884c	rpfC	PROBABLE RESUSCITATION-PROMOTING FACTOR RPFC	66, 80, 88, 94, 107, 139, 141, 153		SP,	CF 2, CF 3, MEM 7, MEM 8, MEM 9, WCL 15, SOL 16		

Rv1885c		CONSERVED HYPOTHETICAL PROTEIN	44		SP, TM	CF 2, CF 3, MEM 8, CW 14		
Rv1887		HYPOTHETICAL PROTEIN	280, 331			CF 3, MEM 7	PhoA 18, BlaTEM 20	
Rv1891		CONSERVED HYPOTHETICAL PROTEIN	58, 77, 102		Tat SP	CF 2, CF 3, MEM 7, MEM 8	BlaTEM 20	
Rv1902c	nanT	PROBABLE SIALIC ACID-TRANSPORT INTEGRAL MEMBRANE PROTEIN NANT	246, 253, 259		TM			Macrophage 24
Rv1903		PROBABLE CONSERVED MEMBRANE PROTEIN	28, 84	Yes	TM	MEM 9		
Rv1906c		CONSERVED HYPOTHETICAL PROTEIN	29, 62, 123	Yes	SP, TM	CF 2, CF 3, MEM 7	PhoA 17	
Rv1911c	lppC	PROBABLE LIPOPROTEIN LPPC	25, 27, 51, 88, 189	Yes	Lipo	CF 2, CF 3, MEM 7, MEM 10, CW 12, WCL 15	PhoA 18	Mouse 27
Rv1921c	lppF	PROBABLE CONSERVED LIPOPROTEIN LPPF	30	Yes	SP, Lipo	MEM 9		
Rv1922		PROBABLE CONSERVED LIPOPROTEIN	25		SP, Lipo	CF 2, MEM 9, CW 14		
Rv1926c	mpt63, mpb63	IMMUNOGENIC PROTEIN MPT63 (ANTIGEN MPT63/MPB63) (16 kDa IMMUNOPROTECTIVE EXTRACELLULAR PROTEIN)	33, 66, 85, 90, 99, 118, 129, 137	Yes	SP, TM	CF 1, CF 2, CF 3, MEM 7, MEM 8, MEM 9, CW 12, CW 13, CW 14, WCL 15, SOL 16	PhoA 18	Mouse 27
Rv1965	yrbE3B	CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBE3B	90, 104		TM			Macaque 22
Rv1968	mce3C	MCE-FAMILY PROTEIN MCE3C	42, 48, 49, 53, 60	Yes	TM			
Rv1969	mce3D	MCE-FAMILY PROTEIN MCE3D	27, 30, 110		SP, TM	CW 12		
Rv1973		POSSIBLE CONSERVED MCE ASSOCIATED MEMBRANE PROTEIN	35, 51	Yes	TM			
Rv1979c		POSSIBLE CONSERVED PERMEASE	374		TM			
Rv1980c	mpt64, mpb64	IMMUNOGENIC PROTEIN MPT64 (ANTIGEN MPT64/MPB64)	22, 36, 119, 210		SP, TM	CF 1, CF 2, CF 3, MEM 7, MEM 8, CW 12, CW 13, CW 14, WCL 15, SOL 16		
Rv1984c	cfp21	PROBABLE CUTINASE PRECURSOR CFP21	23, 24, 25, 28, 43, 67, 132, 157, 184, 186, 215	Yes	SP,	CF 1, CF 2, CF 3, MEM 7, MEM 8, CW 12, CW 13, CW 14	BlaTEM 20	
Rv1986		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	74, 76	Yes	TM			
Rv1987		POSSIBLE CHITINASE	51, 110, 116, 123		SP, TM	CF 2, MEM 8, MEM 9		
Rv1997	ctpF	PROBABLE METAL CATION TRANSPORTER P-TYPE ATPASE A CTPF	278, 318	Yes	TM	MEM 5, MEM 9, MEM 10, CW 13, CW 14		
Rv1999c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	172, 399, 400		Tat SP, TM			
Rv2025c		POSSIBLE CONSERVED MEMBRANE PROTEIN	70, 127		TM			
Rv2046	lppI	Probable lipoprotein lppI	20, 34, 45, 65, 103, 128, 142	Yes	SP, Lipo	MEM 4, MEM 8, MEM 9		Macaque 22
Rv2051c	ppm1	Polyprenol-monophosphomannose	93, 306	Yes	TM	MEM 4, MEM 8, MEM 9, CW 11,		Mouse 27, in vitro 29

		synthase Ppm1				CW 14		
Rv2053c	fxsA	PROBABLE TRANSMEMBRANE PROTEIN	106		TM	MEM 9		
Rv2060		Possible conserved integral membrane protein	70		SP, TM	MEM 7		
Rv2080	lppJ	Possible lipoprotein lppJ	37, 47, 69		SP, Lipo, TM	CF 2, MEM 7, MEM 8, MEM 9	BlaTEM 20	
Rv2088	pknJ	PROBABLE TRANSMEMBRANE SERINE/THREONINE-PROTEIN KINASE J PKNJ (PROTEIN KINASE J) (STPK J)	399, 435			MEM 8, MEM 9, CW 14		
Rv2091c		Probable membrane protein	161, 176		TM	CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		Macaque 22, Mouse 27
Rv2093c	tatC	Probable Sec-independent protein translocase transmembrane protein tatC	110, 241, 249	Yes	TM	MEM 8		in vitro 28, in vitro 29
Rv2113		Probable integral membrane protein	67, 197	Yes	TM	MEM 5, MEM 8, MEM 9, CW 14	BlaTEM 20	
Rv2120c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	16, 30, 96, 118		SP, TM	MEM 8, MEM 9, MEM 10		
Rv2127	ansP1	Probable L-asparagine permease ansP1	183, 193, 202, 236, 272, 276, 293, 364		TM	MEM 8, MEM 9, CW 14	BlaTEM 20	Macrophage 24
Rv2128		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	48	Yes	TM			
Rv2138	lppL	Probable conserved lipoprotein LppL	35		SP, Lipo, TM	CF 2, MEM 8, MEM 9, CW 14		Mouse 27, in vitro 28, in vitro 29
Rv2144c		Probable transmembrane protein	86		TM			
Rv2151c	ftsQ	POSSIBLE CELL DIVISION PROTEIN FTSQ	144, 230		TM	CF 3, MEM 5, MEM 9, MEM 10, CW 13, CW 14		in vitro 28, in vitro 29
Rv2154c	ftsW	FtsW-like protein FtsW	149, 273, 388		TM	CW 14		in vitro 28, in vitro 29
Rv2156c	murX	Probable phospho-N-acetylmuramoyl-pentappeptidetransferase MurX	46, 191, 237		TM	MEM 9, MEM 10, CW 13		in vitro 28, in vitro 29
Rv2171	lppM	Probable conserved lipoprotein lppM	22, 40, 45, 104, 105, 143	Yes	SP, Lipo, TM	CF 2, MEM 8, MEM 9, CW 12, CW 14		
Rv2174		Possible conserved integral membrane protein	175, 199, 263, 290, 339, 347		TM	MEM 8, MEM 9		in vitro 28, in vitro 29
Rv2181		Probable conserved integral membrane protein	64, 285, 401, 417		TM	MEM 8, MEM 9		
Rv2190c		CONSERVED HYPOTHETICAL PROTEIN	27, 107, 174, 229, 238, 258, 269	Yes	SP, TM	MEM 7, MEM 9, CW 14		in vitro 29
Rv2193	ctaE	PROBABLE CYTOCHROME C OXIDASE (SUBUNIT III) CTAE	53, 143, 151, 152	Yes	TM	MEM 9		in vitro 28, in vitro 29
Rv2194	qcrC	Probable Ubiquinol-cytochrome C reductase QcrC(cytochrome C subunit)	111, 139, 180, 197, 243	Yes	SP, TM	CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		in vitro 28, in vitro 29
Rv2198c	mmpS3	PROBABLE CONSERVED MEMBRANE PROTEIN	130, 208, 218, 235		TM	CF 2, CF 3, MEM 5, MEM 8, MEM		Mouse 27, in vitro growth-

		MMPS3				9, MEM 10, CW 12, CW 13, CW 14		defect 28, in vitro 29
Rv2199c	ctaF	Possible conserved integral membrane protein	30, 44, 45, 115, 125	Yes	TM	MEM 10		
Rv2200c	ctaC	PROBABLE TRANSMEMBRANE CYTOCHROME C OXIDASE (SUBUNIT II) CTAC	92, 287, 353		TM	CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15	BlaTEM 20	in vitro 28, in vitro 29
Rv2203		POSSIBLE CONSERVED MEMBRANE PROTEIN	114, 137, 195, 225	Yes	TM	MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15	BlaTEM 20	
Rv2209		Probable conserved integral membrane protein	110, 187		TM	MEM 9		
Rv2219		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	68		TM	MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14		in vitro 28, in vitro 29
Rv2223c		Probable exported protease	31, 37, 39, 84	Yes	SP, TM	CF 2, MEM 9, CW 12	PhoA 17	
Rv2224c		Probable exported protease	64, 65, 110, 113, 141, 325, 360, 383, 478	Yes	SP, Lipo, TM	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 13, CW 14		Macrophage 23, Mouse 25, Mouse 26, Mouse 27
Rv2235		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	80, 112, 145, 198, 210		TM	MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv2240c		HYPOTHETICAL PROTEIN	128, 177, 185, 262	Yes		CF 2, CF 3, MEM 8, MEM 9, CW 14	BlaTEM 20	
Rv2262c		CONSERVED HYPOTHETICAL PROTEIN	88, 89, 227, 239, 244		Tat SP, TM			
Rv2264c		conserved hypothetical proline rich protein	497, 535			CW 14	BlaTEM 20	
Rv2265		Possible conserved integral membrane protein	180, 188, 189, 191		TM			
Rv2270	lppN	PROBABLE LIPOPROTEIN LPPN	27		SP, Lipo			
Rv2272		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	52, 70, 120		TM	MEM 9, CW 14		
Rv2273		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	102		TM			
Rv2281	pitB	Putative phosphate-transport permease PitB	62, 131, 137, 195		TM	MEM 9		
Rv2284	lipM	Probable esterase LipM	62, 105		TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14	BlaTEM 20	
Rv2287	yjcE	Probable conserved integral membrane transport protein YjcE	122, 196, 301	Yes	TM	MEM 9, CW 14		
Rv2289	cdh	Probable CDP-diacylglycerol pyrophosphatase Cdh (CDP-diacylglycerol diphosphatase) (CDP-diacylglycerol phosphatidylhydrolase)	99, 115, 121		SP, TM	MEM 4, MEM 5, MEM 8, MEM 9, CW 14	PhoA 18	
Rv2290	lppO	Probable conserved lipoprotein lppO	44, 46, 59		SP, Lipo	CF 2, CF 3, MEM 9, CW 14	PhoA 17, BlaTEM 20	
Rv2300c		CONSERVED HYPOTHETICAL PROTEIN	210			MEM 8, MEM 9		
Rv2301	cut2, cfp25	PROBABLE CUTINASE CUT2	30, 31, 34, 59, 118, 119, 147, 174, 208		SP, Tat SP, TM	CF 1, CF 2, CF 3, MEM 7, MEM 8, CW 14	BlaTEM 20	
Rv2307c		CONSERVED	49, 50	Yes	SP,	CW 12, CW 14		

		HYPOTHETICAL PROTEIN			TM			
Rv2316	uspA	PROBABLE SUGAR-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER USPA	71, 231, 252		Tat SP, TM			
Rv2317	uspB	PROBABLE SUGAR-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER USPB	36, 63		TM			Mouse 27, in vitro 29
Rv2318	uspC	PROBABLE PERIPLASMIC SUGAR-BINDING LIPOPROTEIN USPC	30, 35, 133		SP, TM			in vitro 29
Rv2320c	rocE	PROBABLE CATIONIC AMINO ACID TRANSPORT INTEGRAL MEMBRANE PROTEIN ROCE	227, 231, 296, 297, 395, 450	Yes	TM	MEM 8	BlaTEM 20	
Rv2325c		CONSERVED HYPOTHETICAL PROTEIN	67, 100, 109, 125	Yes	TM	MEM 4, MEM 9, CW 14		in vitro 28
Rv2326c		POSSIBLE TRANSMEMBRANE ATP-BINDING PROTEIN ABC TRANSORTER	59, 69, 99, 114, 147	Yes	TM	MEM 4, MEM 5, MEM 8, MEM 9, CW 12, CW 14		
Rv2329c	narK1	PROBABLE NITRITE EXTRUSION PROTEIN 1 NARK1 (NITRITE FACILITATOR 1)	154, 251, 380, 446	Yes	TM			
Rv2330c	lppP	PROBABLE LIPOPROTEIN LPPP	28, 52		SP, Lipo, TM	MEM 8, MEM 9, CW 14		Macrophage 23
Rv2333c	stp	PROBABLE CONSERVED INTEGRAL MEMBRANE TRANSPORT PROTEIN	99, 290, 297, 388, 450, 503	Yes	TM			
Rv2339	mmpL 9	PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL9	72, 193, 202, 406, 422, 877	Yes	TM	MEM 8, MEM 9, CW 12, CW 14	BlaTEM 20	in vitro 29
Rv2345		POSSIBLE CONSERVED TRANSMEMBRANE PROTEIN	15, 22, 45, 48, 74, 103	Yes	SP, TM	CF 2, MEM 4, MEM 5, MEM 6, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		
Rv2376c	cfp2, mtb12	LOW MOLECULAR WEIGHT ANTIGEN CFP2 (LOW MOLECULAR WEIGHT PROTEIN ANTIGEN 2) (CFP-2)	21, 29, 51		SP, TM	CF 1, CF 2, CF 3, MEM 7, MEM 8, CW 12, CW 13, WCL 15, SOL 16		
Rv2380c	mbtE	PEPTIDE SYNTHETASE MBTE (PEPTIDE SYNTHASE)				MEM 4, MEM 9, CW 13, SOL 16		Mouse 27, in vitro 29
Rv2387		CONSERVED HYPOTHETICAL PROTEIN	55, 146	Yes	TM	MEM 8, MEM 9, CW 14		Macaque 22, Mouse 25
Rv2389c	rpfD	PROBABLE RESUSCITATION-PROMOTING FACTOR RPFD	61		SP, TM			
Rv2390c		CONSERVED HYPOTHETICAL PROTEIN	45, 50, 64	Yes	TM			
Rv2394	ggtB	PROBABLE GAMMA-GLUTAMYLTRANSPEPTIDASE PRECURSOR GGTB (GAMMA-GLUTAMYLTRANSFERASE) (GLUTAMYL TRANSPEPTIDASE)	18, 21, 31, 65, 121, 124, 126, 151, 213, 259, 266, 270, 271, 278, 396, 525, 578	Yes	SP, Tat SP, Lipo	CF 2, CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14	BlaTEM 20	
Rv2395		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	95, 342, 346, 347, 485, 515		TM			

Rv2399c	cysT	PROBABLE SULFATE-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER CYST	52, 143, 145		TM	MEM 9, CW 14		in vitro 28, in vitro 29
Rv2400c	subI	PROBABLE SULFATE-BINDING LIPOPROTEIN SUBI	74, 99, 218, 232, 354		Lipo	CF 2, CF 3, MEM 5, MEM 8, MEM 9, CW 14		Mouse 27, in vitro 28, in vitro 29
Rv2403c	lppR	PROBABLE CONSERVED LIPOPROTEIN LPPR	24		SP, Lipo, TM	MEM 9, CW 14		
Rv2434c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	143		TM			
Rv2435c		PROBABLE CYCLASE (ADENYLYL-OR GUANYLYL-)(ADENYLATE-OR GUANYLATE-)	69, 148, 155	Yes	TM	MEM 9		
Rv2437		CONSERVED HYPOTHETICAL PROTEIN	91		TM			Mouse 25, in vitro 28
Rv2443	dctA	PROBABLE C4-DICARBOXYLATE-TRANSPORT TRANSMEMBRANE PROTEIN DCTA	46, 130, 135		TM	MEM 8, MEM 9	BlaTEM 20	
Rv2450c	rpfE	PROBABLE RESUSCITATION-PROMOTING FACTOR RPFE	22, 38, 108, 116, 127, 128	Yes	SP,	MEM 7	PhoA 17	
Rv2459		PROBABLE CONSERVED INTEGRAL MEMBRANE TRANSPORT PROTEIN	222, 306		TM			Macaque 22
Rv2473		POSSIBLE ALANINE AND PROLINE RICH MEMBRANE PROTEIN	115, 133	Yes	SP, TM	MEM 8, MEM 9, CW 14		
Rv2507		POSSIBLE CONSERVED PROLINE RICH MEMBRANE PROTEIN	148, 167, 175, 183, 207, 234	Yes	TM	MEM 8		in vitro 28, in vitro 29
Rv2508c		PROBABLE CONSERVED INTEGRAL MEMBRANE LEUCINE AND ALANINE RICH PROTEIN	117, 179, 224, 232, 297, 348	Yes	TM	MEM 8, MEM 9		
Rv2518c	lppS	PROBABLE CONSERVED LIPOPROTEIN LPPS	34, 38, 81, 157, 402	Yes	SP, Lipo, TM	MEM 8, MEM 9, CW 12, CW 14		Mouse 27, in vitro 29
Rv2519	PE26	PE FAMILY PROTEIN	255, 333		SP,			
Rv2536		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	32, 82, 86, 102, 116		TM	CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		Mouse 27
Rv2543	lppA	PROBABLE CONSERVED LIPOPROTEIN LPPA	45		Lipo			Mouse 27
Rv2544	lppB	PROBABLE CONSERVED LIPOPROTEIN LPPB	45		Lipo	MEM 7		Mouse 27
Rv2553c		PROBABLE CONSERVED MEMBRANE PROTEIN	76, 78, 89, 127	Yes	TM	CF 3, MEM 9, CW 14		Mouse 25, Mouse 27, in vitro 29
Rv2560		PROBABLE PROLINE AND GLYCINE RICH TRANSMEMBRANE PROTEIN	135, 170, 174		TM	MEM 8, MEM 9		
Rv2563		PROBABLE GLUTAMINE-TRANSPORT TRANSMEMBRANE PROTEIN ABC TRANSPORTER	33, 59, 88, 108, 117, 165	Yes	TM	MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 13, CW 14		Mouse 27
Rv2571c		PROBABLE TRANSMEMBRANE	94	Yes	TM	MEM 9, CW 14		

		ALANINE AND VALINE AND LEUCINE RICH PROTEIN						
Rv2575		POSSIBLE CONSERVED MEMBRANE GLYCINE RICH PROTEIN	46, 63, 72, 100, 116, 154, 287	Yes	Tat SP, TM	CF 2, CF 3, MEM 7, MEM 8, CW 14, SOL 16		
Rv2576c		POSSIBLE CONSERVED MEMBRANE PROTEIN	95, 102, 115, 117, 144	Yes	SP, TM	CF 2, CF 3, MEM 7, CW 12		
Rv2582	ppiB, ppi	PROBABLE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE B PPIB (CYCLOPHILIN) (PPIASE) (ROTAMASE) (PEPTIDYLPROLYL ISOMERASE)	93, 95, 130, 204, 234, 242, 284, 287	Yes	TM	CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		Mouse 27, in vitro 28, in vitro 29
Rv2586c	secF	PROBABLE PROTEIN-EXPORT MEMBRANE PROTEIN SECF	94, 130	Yes	TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 14		in vitro 29
Rv2587c	secD	PROBABLE PROTEIN-EXPORT MEMBRANE PROTEIN SECD	30, 35, 45, 52, 83, 99, 144, 268, 304, 348	Yes	TM	MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15	PhoA 18	in vitro 29
Rv2599		PROBABLE CONSERVED MEMBRANE PROTEIN	108, 129, 136		SP, TM	CF 2, MEM 8, MEM 9, CW 14	BlaTEM 20	
Rv2615c	PE_PGRS45	PE-PGRS FAMILY PROTEIN	303		SP,			
Rv2617c		PROBABLE TRANSMEMBRANE PROTEIN	109, 111		TM	MEM 8, MEM 9, CW 14		
Rv2639c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	62, 70, 107		TM		BlaTEM 20	
Rv2643	arsC	PROBABLE ARSENIC-TRANSPORT INTEGRAL MEMBRANE PROTEIN ARSC	121, 256, 326		TM	CW 14		
Rv2668		POSSIBLE EXPORTED ALANINE AND VALINE RICH PROTEIN	28, 44, 70, 107, 129		SP,	CF 2, CF 3, MEM 7		
Rv2672		POSSIBLE SECRETED PROTEASE	89, 135, 209, 223, 230, 327, 330, 334, 369, 373, 526	Yes	SP, Lipo, TM	CF 2, CF 3, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 13, CW 14		Mouse 27
Rv2684	arsA	PROBABLE ARSENIC-TRANSPORT INTEGRAL MEMBRANE PROTEIN ARSA	50		TM	MEM 9		
Rv2686c		PROBABLE ANTIBIOTIC-TRANSPORT INTEGRAL MEMBRANE LEUCINE AND ALANINE AND VALINE RICH PROTEIN ABC TRANSPORTER	135		TM			Macaque 22
Rv2690c		PROBABLE CONSERVED INTEGRAL MEMBRANE ALANINE AND VALINE AND LEUCINE RICH PROTEIN	205, 283, 394, 463	Yes	TM	MEM 9, CW 14		Macaque 22, Mouse 27, in vitro 29
Rv2698		PROBABLE CONSERVED ALANINE RICH TRANSMEMBRANE PROTEIN	58	Yes	TM	MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv2700		POSSIBLE CONSERVED SECRETED ALANINE RICH PROTEIN	84, 120, 128, 196, 204		TM	CF 3, MEM 5, MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv2719c		POSSIBLE CONSERVED	93, 94,	Yes	TM			

		MEMBRANE PROTEIN	104					
Rv2721c		POSSIBLE CONSERVED TRANSMEMBRANE ALANINE AND GLYCINE RICH PROTEIN	46, 77, 78, 87, 104, 133, 152, 222, 235, 250, 269, 315, 334, 339, 370, 387, 395, 413	Yes	SP, TM	CF 2, CF 3, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14	BlaTEM 20	
Rv2723		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	59, 122, 231, 287, 313		TM	MEM 8		
Rv2729c		PROBABLE CONSERVED INTEGRAL MEMBRANE ALANINE VALINE AND LEUCINE RICH PROTEIN	127, 226, 287, 288		TM			
Rv2732c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	53, 57, 68, 130	Yes	Tat SP, TM	MEM 8, MEM 9, CW 14		
Rv2746c	pgsA3	PROBABLE PGP SYNTHASE PGSA3 (CDP-DIACYLGLYCEROL--GLYCEROL-3-PHOSPHATE 3-PHOSPHATIDYLTRANSFERASE) (PHOSPHATIDYLGLYCEROPHOSPHATE SYNTHASE)	50, 55, 67, 179	Yes	TM	MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv2748c	ftsK	POSSIBLE CELL DIVISION TRANSMEMBRANE PROTEIN FTSK	168, 246	Yes	TM	MEM 9, CW 11, CW 12, CW 14	PhoA 18	in vitro 28, in vitro 29
Rv2772c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	41, 64		Tat SP, TM	MEM 8, MEM 9, CW 14		
Rv2784c	lppU	PROBABLE LIPOPROTEIN LPPU	165, 166		SP, Lipo	MEM 8, MEM 9		
Rv2796c	lppV	PROBABLE CONSERVED LIPOPROTEIN LPPV	18, 27, 86		SP, Lipo	MEM 9, CW 14		Macaque 22
Rv2799		PROBABLE MEMBRANE PROTEIN	47, 48, 57, 140, 144		TM	CF 2, CF 3, MEM 7, CW 14		
Rv2806		POSSIBLE MEMBRANE PROTEIN	38		TM			
Rv2835c	ugpA	PROBABLE Sn-GLYCEROL-3-PHOSPHATE TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER UGPA	247, 271		TM			
Rv2843		PROBABLE CONSERVED TRANSMEMBRANE ALANINE RICH PROTEIN	44, 65		SP, Tat SP, Lipo, TM	MEM 8, MEM 9, CW 14	BlaC 19	
Rv2846c	efpA	POSSIBLE INTEGRAL MEMBRANE EFFLUX PROTEIN EFPA	87, 94, 160, 186		TM	MEM 8, MEM 9		in vitro 28, in vitro 29
Rv2856	nicT	POSSIBLE NICKEL-TRANSPORT INTEGRAL MEMBRANE PROTEIN NICT	51, 57, 152, 328	Yes	TM			in vitro 28
Rv2864c		POSSIBLE PENICILLIN-BINDING LIPOPROTEIN	31, 40, 47, 49	Yes	SP, Lipo	MEM 9, CW 12		
Rv2869c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	155, 165, 223, 238, 250, 259, 285, 392		TM	CF 2, MEM 8, MEM 9, CW 14		Mouse 27, in vitro 28, in vitro 29
Rv2873	mpt83, mpb83	CELL SURFACE LIPOPROTEIN MPT83 (LIPOPROTEIN P23)	54, 118, 198		SP, Lipo	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM		

						9, MEM 10, CW 13, CW 14, WCL 15		
Rv2874	dipZ	POSSIBLE INTEGRAL MEMBRANE C-TYPE CYTOCHROME BIOGENESIS PROTEIN DIPZ	143, 353, 362, 396, 406, 412, 545, 583	Yes	TM	MEM 7		
Rv2877c	, merT	PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	75, 158		TM	MEM 8, MEM 9		
Rv2903c	lepB	PROBABLE SIGNAL PEPTIDASE I LEPB (SPASE I) (LEADER PEPTIDASE I).	179, 229		TM	MEM 5, MEM 8, MEM 9, CW 14	PhoA 18	in vitro 28, in vitro 29
Rv2905	lppW	PROBABLE CONSERVED ALANINE RICH LIPOPROTEIN LPPW	85, 115, 166, 212		SP, Lipo	MEM 7, CW 14	BlaTEM 20	
Rv2911	dacB2, dacB	PROBABLE D-ALANYL-D-ALANINE CARBOXYPEPTIDASE DACB2 (PENICILLIN-BINDING PROTEIN) (DD-PEPTIDASE) (DD-CARBOXYPEPTIDASE) (PBP) (DD-TRANSPEPTIDASE) (SERINE-TYPE D-ALA-D-ALA CARBOXYPEPTIDASE) (D-AMINO ACID HYDROLASE)	21, 31, 51, 56, 66, 127, 168, 169, 171, 243, 246	Yes	SP, Tat SP	CF 2, CF 3, MEM 7, CW 13, CW 14, SOL 16		Macaque 22
Rv2914c	pknI	PROBABLE TRANSMEMBRANE SERINE/THREONINE-PROTEIN KINASE I PKNI (PROTEIN KINASE I) (STPK I) (PHOSPHORYLASE B KINASE KINASE) (HYDROXYALKYL-PROTEIN KINASE)	376, 582	Yes	TM	MEM 9, CW 14		
Rv2920c	amt	PROBABLE AMMONIUM-TRANSPORT INTEGRAL MEMBRANE PROTEIN AMT	79, 105, 175, 179, 192, 195, 201, 257, 382, 428	Yes	TM	MEM 8		Macaque 22
Rv2938	drrC	PROBABLE DAUNORUBICIN-DIM-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER DRRC	235		TM	MEM 8, MEM 9, CW 12, CW 14	BlaTEM 20	Mouse 25
Rv2942	mmpL 7	CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN MMPL7	95, 276, 420, 438, 487	Yes	TM	MEM 5, MEM 8, MEM 9, MEM 10, CW 14		Macrophage 23, Mouse 25
Rv2945c	lppX	PROBABLE CONSERVED LIPOPROTEIN LPPX	47	Yes	SP, Lipo, TM	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16	PhoA 18, BlaTEM 20	Mouse 25, Mouse 27
Rv2963		PROBABLE INTEGRAL MEMBRANE PROTEIN	127, 347		TM	MEM 8, MEM 9		
Rv2968c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	64		TM	MEM 6		in vitro 28, in vitro 29
Rv2969c		POSSIBLE CONSERVED MEMBRANE OR SECRETED PROTEIN	56, 71, 91, 96	Yes	TM	CF 2, CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM		in vitro 28, in vitro 29

						10, CW 12, CW 13, CW 14, WCL 15		
Rv2972c		POSSIBLE CONSERVED MEMBRANE OR EXPORTED PROTEIN	35, 45, 51	Yes	SP, Tat SP, TM	MEM 9		
Rv2994		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	42, 113, 114, 120, 248, 331, 392	Yes	TM	MEM 7, MEM 8, MEM 9, CW 14		
Rv2999	lppY	PROBABLE CONSERVED LIPOPROTEIN LPPY	36, 68, 95, 160, 194, 246	Yes	SP, Lipo	MEM 4, MEM 5, MEM 8, MEM 9, CW 14		in vitro 28
Rv3005c		CONSERVED HYPOTHETICAL PROTEIN	166, 171	Yes	TM	MEM 9, CW 14		Mouse 27
Rv3006	lppZ	PROBABLE CONSERVED LIPOPROTEIN LPPZ	21, 95, 118, 198, 232, 343	Yes	SP, Lipo	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15		in vitro 29
Rv3016	lpqA	PROBABLE LIPOPROTEIN LPQA	24, 32, 55, 57, 129		SP, Lipo	CF 2, MEM 7, CW 14		
Rv3033		HYPOTHETICAL PROTEIN	46, 52, 53, 167		SP,	CF 2, CF 3, MEM 4, MEM 7, MEM 8, MEM 9, CW 13, CW 14		Macrophage 23
Rv3035		CONSERVED HYPOTHETICAL PROTEIN	2, 9, 20, 44, 48, 133, 149, 224, 254, 259, 285, 307, 328	Yes		MEM 8, MEM 9, CW 14		in vitro 29
Rv3036c	TB22.2	PROBABLE CONSERVED SECRETED PROTEIN TB22.2	27, 30, 35, 100, 148, 168, 172, 178	Yes	SP,	CF 1, CF 2, CF 3, MEM 7, CW 14, WCL 15, SOL 16	BlaTEM 20	Mouse 27
Rv3043c	ctaD	PROBABLE CYTOCHROME C OXIDASE POLYPEPTIDE I CTAD (CYTOCHROME AA3 SUBUNIT 1)	66, 68, 137, 143, 165, 371		TM	MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		in vitro 28, in vitro 29
Rv3044	fecB	PROBABLE FEIII-DICITRATE-BINDING PERIPLASMIC LIPOPROTEIN FECB	40, 60, 113, 135, 179, 195, 208, 276		SP, Lipo	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 9, MEM 10, CW 13, CW 14, WCL 15		Mouse 27, in vitro 29
Rv3063	cstA	PROBABLE CARBON STARVATION PROTEIN A HOMOLOG CSTA	269, 348, 357, 441, 464, 477, 478, 578	Yes	Tat SP, TM			
Rv3064c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	52, 56		TM			
Rv3067		CONSERVED HYPOTHETICAL PROTEIN	29, 41, 44, 67, 76	Yes		CF 2		
Rv3090		HYPOTHETICAL ALANINE AND VALINE RICH PROTEIN	71, 119, 133, 166, 188, 197		TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, SOL 16	BlaTEM 20	
Rv3092c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	113, 123, 134, 144		TM	MEM 8, MEM 9, CW 14		
Rv3103c		HYPOTHETICAL PROLINE-RICH PROTEIN	127		TM		BlaTEM 20	Mouse 25
Rv3123		HYPOTHETICAL PROTEIN	145			MEM 9		
Rv3152	nuoH	PROBABLE NADH DEHYDROGENASE I (CHAIN H) NUOH (NADH-	280, 289	Yes	Tat SP, TM	MEM 8, MEM 9, CW 14		

		UBIQUINONE OXIDOREDUCTASE CHAIN H)						
Rv3156	nuoL	PROBABLE NADH DEHYDROGENASE I (CHAIN L) NUOL (NADH-UBIQUINONE OXIDOREDUCTASE CHAIN L)	75, 76, 77, 211, 233, 354, 414, 422		TM	MEM 8, MEM 9, CW 14		
Rv3157	nuoM	PROBABLE NADH DEHYDROGENASE I (CHAIN M) NUOK (NADH-UBIQUINONE OXIDOREDUCTASE CHAIN M)	74, 218, 233, 530	Yes	SP, Tat SP, TM	MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		in vitro 29
Rv3158	nuoN	PROBABLE NADH DEHYDROGENASE I (CHAIN N) NUON (NADH-UBIQUINONE OXIDOREDUCTASE CHAIN N)	368, 447, 450, 451		TM	MEM 5, MEM 7, MEM 8, MEM 9, CW 14		
Rv3165c		HYPOTHETICAL PROTEIN	29	Yes	TM	MEM 8, MEM 9, CW 14		
Rv3166c		CONSERVED HYPOTHETICAL PROTEIN	29, 43, 44, 58, 137	Yes	TM	MEM 9, CW 14		Mouse 27
Rv3193c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	142, 239, 300, 328, 451, 529, 531, 678	Yes	TM	CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		Mouse 27, in vitro 29
Rv3194c		POSSIBLE CONSERVED SECRETED PROTEIN	33, 111, 119, 152, 171, 194, 238, 264	Yes	SP, TM	CF 3, MEM 5, MEM 7, MEM 8, MEM 9, CW 14		Mouse 27
Rv3207c		CONSERVED HYPOTHETICAL PROTEIN	33, 50, 58, 120, 152, 204		SP, TM	CF 2, MEM 9, CW 14		Mouse 27
Rv3209		CONSERVED HYPOTHETICAL THREONIN AND PROLINE RICH PROTEIN	70, 81	Yes	SP,		BlaTEM 20	Mouse 27
Rv3217c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	47, 55, 60, 61	Yes	TM			
Rv3236c	, kefB	PROBABLE CONSERVED INTEGRAL MEMBRANE TRANSPORT PROTEIN	51, 53, 59		TM	MEM 8, CW 14		Mouse 25
Rv3239c		PROBABLE CONSERVED TRANSMEMBRANE TRANSPORT PROTEIN	446		Tat SP, TM			
Rv3244c	lpqB	PROBABLE CONSERVED LIPOPROTEIN LPQB	174, 232, 336		SP, Lipo	CF 2, CF 3, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 14, SOL 16		in vitro 28, in vitro 29
Rv3245c	mtrB	TWO COMPONENT SENSORY TRANSDUCTION HISTIDINE KINASE MTRB	88, 116, 154, 175		TM	MEM 9, CW 14		Mouse 27, in vitro 28, in vitro 29
Rv3252c	alkB	PROBABLE TRANSMEMBRANE ALKANE 1-MONOOXYGENASE ALKB (ALKANE 1-HYDROXYLASE) (LAURIC ACID OMEGA-HYDROXYLASE) (OMEGA-HYDROXYLASE) (FATTY ACID OMEGA-	55		TM	MEM 9		

		HYDROXYLASE) (ALKANE HYDROXYLASE-RUBREDOXIN)						
Rv3253c		POSSIBLE CATIONIC AMINO ACID TRANSPORT INTEGRAL MEMBRANE PROTEIN	49, 52, 199, 215, 227, 241, 314, 325, 334, 413, 453, 468	Yes	TM	MEM 9	BlaTEM 20	
Rv3267		CONSERVED HYPOTHETICAL PROTEIN (CPSA-RELATED PROTEIN)	28, 87	Yes	SP, TM	CF 2, CF 3, MEM 7, MEM 9, MEM 10, CW 13, CW 14	PhoA 18, BlaTEM 20	Mouse 27, in vitro 28, in vitro 29
Rv3271c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	62, 68, 83, 89, 120, 203	Yes	TM	MEM 9, CW 14		in vitro 28
Rv3273		PROBABLE TRANSMEMBRANE CARBONIC ANHYDRASE (CARBONATE DEHYDRATASE) (CARBONIC DEHYDRATASE)	55, 61, 69, 73, 366	Yes	TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		
Rv3274c	fadE25	PROBABLE ACYL-CoA DEHYDROGENASE FADE25	316			CF 2, CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		Mouse 27
Rv3278c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	42	Yes	TM	MEM 8, MEM 9, CW 14		
Rv3289c		POSSIBLE TRANSMEMBRANE PROTEIN	52		TM			
Rv3298c	lpqC	POSSIBLE ESTERASE LIPOPROTEIN LPQC	19, 49, 53, 63, 65, 67, 81, 120, 131, 137, 252	Yes	SP, Lipo	MEM 5, MEM 8, MEM 9, CW 14		
Rv3310		POSSIBLE ACID PHOSPHATASE (ACID PHOSPHOMONOESTERASE) (PHOSPHOMONOESTERASE) (GLYCEROPHOSPHATASE)	78, 118, 131, 186, 258		SP, TM	CF 2, CF 3, MEM 7, MEM 8		
Rv3312A	mtp	SECRETED PROTEIN ANTIGEN	32, 48, 77		SP, TM	CF 3, WCL 15, SOL 16		
Rv3316	sdhC	PROBABLE SUCCINATE DEHYDROGENASE (CYTOCHROME B-556 SUBUNIT) SDHC (SUCCINIC DEHYDROGENASE) (FUMARATE REDUCTASE) (FUMARATE DEHYDROGENASE) (FUMARIC HYDROGENASE)	24		TM	MEM 8, MEM 9		Mouse 27
Rv3330	dacB1	PROBABLE PENICILLIN-BINDING PROTEIN DACB1 (D-ALANYL-D-ALANINE CARBOXYPEPTIDASE) (DD-PEPTIDASE) (DD-CARBOXYPEPTIDASE) (PBP) (DD-TRANSPEPTIDASE) (SERINE-TYPE D-ALA-D-	24, 26, 31, 52, 54, 69, 76, 80, 88, 108, 141, 158, 192, 215	Yes	SP, TM	CF 2, MEM 9, CW 14		

		ALA CARBOXYPEPTIDASE) (D-AMINO ACID HYDROLASE)						
Rv3331	sugI	PROBABLE SUGAR-TRANSPORT INTEGRAL MEMBRANE PROTEIN SUGI	200, 212, 312, 332	Yes	TM	MEM 9, CW 14		
Rv3343c	PPE54	PPE FAMILY PROTEIN	555					in vitro 28, in vitro 29
Rv3350c	PPE56	PPE FAMILY PROTEIN	3387					in vitro 29
Rv3354		CONSERVED HYPOTHETICAL PROTEIN	26, 40, 45, 113, 126		SP, TM	MEM 7		
Rv3365c		CONSERVED HYPOTHETICAL PROTEIN	79, 104, 116, 145, 203, 215, 232, 370	Yes	TM	MEM 9, CW 14		
Rv3390	lpqD	PROBABLE CONSERVED LIPOPROTEIN LPQD	25, 57, 59, 95, 96, 100, 118, 154, 164	Yes	SP, Lipo, TM	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14	BlaTEM 20	
Rv3395A		PROBABLE MEMBRANE PROTEIN	28, 30, 60, 103	Yes	SP, TM	CF 2		Mouse 27
Rv3413c		HYPOTHETICAL ALANINE AND PROLINE RICH PROTEIN	110, 200, 277	Yes	TM	CF 2, CF 3, MEM 7	BlaTEM 20	
Rv3434c		POSSIBLE CONSERVED TRANSMEMBRANE PROTEIN	44, 191		TM			Mouse 27
Rv3435c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	27, 106, 116, 119	Yes	TM	MEM 8, MEM 9, CW 14		
Rv3451	cut3	PROBABLE CUTINASE PRECURSOR CUT3	52, 54, 58, 87, 134, 169, 198, 203, 227, 241	Yes	SP, TM	CF 2, MEM 8		
Rv3452	cut4	PROBABLE CUTINASE PRECURSOR CUT4	36, 55, 57, 124, 137, 142, 152, 159, 172	Yes	SP, TM			
Rv3476c	kgtP	PROBABLE DICARBOXYLIC ACID TRANSPORT INTEGRAL MEMBRANE PROTEIN KGTP (DICARBOXYLATE TRANSPORTER)	284		TM	MEM 8	BlaTEM 20	
Rv3478	PPE60, mtb39c	PE FAMILY PROTEIN	155, 177			MEM 8, MEM 9, CW 13, CW 14		
Rv3481c		PROBABLE INTEGRAL MEMBRANE PROTEIN	62, 69, 73, 147, 158		TM	MEM 8, MEM 9		
Rv3482c		PROBABLE CONSERVED MEMBRANE PROTEIN	104, 109		TM	MEM 5, MEM 9, CW 14		
Rv3483c		CONSERVED HYPOTHETICAL PROTEIN	52, 70, 126, 154		TM	MEM 5, MEM 8, MEM 9, CW 14		
Rv3484	cpsA	POSSIBLE CONSERVED PROTEIN CPSA	39, 51, 67, 105, 193, 206, 216, 231, 236, 237, 321, 348, 384, 395, 401, 452, 487	Yes	TM	CF 2, MEM 7, MEM 8	BlaTEM 20	Macrophage 24, Mouse 25, Mouse 27, in vitro 29
Rv3491		HYPOTHETICAL PROTEIN	24, 44, 47, 54, 101		SP, TM	CF 2, MEM 7		Mouse 27
Rv3492c		CONSERVED HYPOTHETICAL MCE ASSOCIATED PROTEIN	19, 21	Yes	SP, TM	MEM 9, CW 14		
Rv3493c		CONSERVED	118, 150,		TM	MEM 5, MEM 9,		Mouse 27

		HYPOTHETICAL MCE ASSOCIATED ALANINE AND VALINE RICH PROTEIN	180, 230			CW 14		
Rv3494c	mce4F	MCE-FAMILY PROTEIN MCE4F	31, 48, 60, 171, 354, 363	Yes	TM	CF 3, MEM 9, CW 14	BlaTEM 20	Macaque 22, Mouse 27
Rv3496c	mce4D	MCE-FAMILY PROTEIN MCE4D	35, 418, 448		TM	MEM 9, CW 12, CW 14	BlaTEM 20	Mouse 27
Rv3497c	mce4C	MCE-FAMILY PROTEIN MCE4C	53, 58, 78, 120, 303, 305, 316	Yes	TM	MEM 5, MEM 8, MEM 9, CW 14	BlaTEM 20	Mouse 25, Mouse 27
Rv3498c	mce4B	MCE-FAMILY PROTEIN MCE4B	27, 63, 137, 302	Yes	SP, TM	MEM 9, CW 14	BlaTEM 20	Mouse 27
Rv3499c	mce4A, mce4	MCE-FAMILY PROTEIN MCE4A	24, 49, 87, 102, 171, 196, 215, 216	Yes	SP, TM	MEM 6, MEM 9, CW 14		Mouse 25
Rv3500c	yrbE4B	CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBE4B	78, 99, 193, 272		TM	MEM 8, MEM 9, CW 14		
Rv3501c	yrbE4A	CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YRBE4A	74, 78, 80, 159, 160, 184, 244	Yes	TM	MEM 8, MEM 9, CW 14		Mouse 25
Rv3507	PE_PGRS53	PE-PGRS FAMILY PROTEIN	820, 1320		SP, Tat SP			
Rv3524		PROBABLE CONSERVED MEMBRANE PROTEIN	122, 132, 149, 174, 248, 257, 274, 275, 300, 311		TM	MEM 9, CW 14		
Rv3526		POSSIBLE OXIDOREDUCTASE	381					
Rv3554	fdxB	POSSIBLE ELECTRON TRANSFER PROTEIN FDXB	65		TM	MEM 9, CW 14		
Rv3572		HYPOTHETICAL PROTEIN	24, 74, 129, 139	Yes	SP,	CF 2, CF 3, MEM 7, MEM 8, CW 14		
Rv3576	lppH, pknM	POSSIBLE CONSERVED LIPOPROTEIN LPPH	18, 27, 34, 59, 102, 135, 163, 168, 190	Yes	SP, Lipo	CF 3, MEM 8, MEM 9, CW 14		
Rv3578	arsB2	POSSIBLE ARSENICAL PUMP INTEGRAL MEMBRANE PROTEIN ARSB2	293, 297, 302, 307		SP, TM			Mouse 27
Rv3584	lpqE	POSSIBLE CONSERVED LIPOPROTEIN LPQE	27, 106, 112, 114, 163	Yes	SP, Lipo, TM	CF 2, CF 3, MEM 4, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15	BlaTEM 20	
Rv3587c		PROBABLE CONSERVED MEMBRANE PROTEIN	44, 95, 156	Yes	SP, TM	CF 2, CF 3, MEM 7, MEM 8, MEM 9, CW 14		in vitro 29
Rv3593	lpqF	PROBABLE CONSERVED LIPOPROTEIN LPQF	38, 82, 104, 186, 231	Yes	SP, Lipo			in vitro 28, in vitro 29
Rv3596c	clpC1, clpC	PROBABLE ATP-DEPENDENT PROTEASE ATP-BINDING SUBUNIT CLPC1	598			CF 3, MEM 4, MEM 6, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		Macrophage 23, Macrophage 24, in vitro 28, in vitro 29
Rv3604c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN RICH IN	33	Yes	TM	MEM 6, MEM 9, CW 14		in vitro 28, in vitro 29

		ALANINE AND ARGININE AND PROLINE						
Rv3610c	ftsH	MEMBRANE-BOUND PROTEASE FTSH (CELL DIVISION PROTEIN)	30, 84		TM	MEM 4, MEM 9, MEM 10, CW 13, CW 14, WCL 15		Mouse 27, in vitro 28, in vitro 29
Rv3623	lpqG	PROBABLE CONSERVED LIPOPROTEIN LPQG	42, 61, 112, 116, 144, 145, 214, 223		SP, Lipo	MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		
Rv3627c		CONSERVED HYPOTHETICAL PROTEIN	27, 62, 68, 72, 96, 139, 140, 163, 200, 228, 260, 335, 381, 401, 408, 438	Yes	SP, TM	CF 2, MEM 5, MEM 7, MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv3629c		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	27, 89, 124, 200, 333	Yes	TM	MEM 7		
Rv3635		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	43, 64, 397, 423		TM	MEM 9		in vitro 28, in vitro 29
Rv3654c		CONSERVED HYPOTHETICAL PROTEIN	16, 18, 23					
Rv3655c		CONSERVED HYPOTHETICAL PROTEIN	42		SP, TM			
Rv3664c	dppC	PROBABLE DIPEPTIDE-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER DPPC	38		SP, TM			Macrophage 24, Mouse 27
Rv3665c	dppB	PROBABLE DIPEPTIDE-TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER DPPB	30, 70, 129		TM			Mouse 27, in vitro 29
Rv3666c	dppA	PROBABLE PERIPLASMIC DIPEPTIDE-BINDING LIPOPROTEIN DPPA	22, 25, 38, 48, 64, 97, 154, 215, 254, 261, 288	Yes	SP, Lipo, TM	CF 2, MEM 8, CW 12, CW 14		Mouse 27, in vitro 28
Rv3667	acs	ACETYL-COENZYME A SYNTHETASE ACS (ACETATE--CoA LIGASE) (ACETYL-CoA SYNTHETASE) (ACETYL-CoA SYNTHASE) (ACYL-ACTIVATING ENZYME) (ACETATE THIOKINASE) (ACETYL-ACTIVATING ENZYME) (ACETATE--COENZYME A LIGASE) (ACETYL-COENZYME A SYNTHASE)				MEM 4, MEM 9, CW 13, CW 14, WCL 15, SOL 16		
Rv3668c		POSSIBLE PROTEASE	25, 35, 92, 144	Yes	SP, TM	CF 2, CF 3, MEM 7	PhoA 17	Mouse 27
Rv3671c		POSSIBLE MEMBRANE-ASSOCIATED SERINE PROTEASE	63, 166, 296, 305, 391	Yes	TM	CF 2, CF 3, MEM 5, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16		Mouse 27, in vitro 29
Rv3673c		POSSIBLE MEMBRANE-ANCHORED THIOREDOXIN-LIKE PROTEIN (THIOL-DISULFIDE INTERCHANGE RELATED PROTEIN)	44, 53, 92		TM	MEM 8, MEM 9, CW 14		in vitro 29
Rv3682	ponA2	PROBABLE	33, 117,	Yes	SP,	CF 2, CF 3, MEM		Macrophage

		BIFUNCTIONAL MEMBRANE-ASSOCIATED PENICILLIN-BINDING PROTEIN 1A/1B PONA2 (MUREIN POLYMERASE) [INCLUDES: PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE (PEPTIDOGLYCAN TGASE) + PENICILLIN-SENSITIVE TRANSPEPTIDASE (DD-TRANSPEPTIDASE)]	124, 130, 285, 290, 317, 335, 346, 379, 395, 405, 432, 466, 597, 613, 734, 746		TM	4, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 14		23, Mouse 27
Rv3691		CONSERVED HYPOTHETICAL PROTEIN	2			MEM 4, MEM 8, MEM 9, CW 14		
Rv3694c		POSSIBLE CONSERVED TRANSMEMBRANE PROTEIN	128, 153, 197		TM	MEM 8, MEM 9, CW 14		
Rv3695		POSSIBLE CONSERVED MEMBRANE PROTEIN	45, 56, 123		TM	MEM 8, MEM 9, CW 14		
Rv3701c		CONSERVED HYPOTHETICAL PROTEIN				MEM 9, CW 14		Macrophage 23, Mouse 25
Rv3705A		CONSERVED HYPOTHETICAL PROLINE RICH PROTEIN	82, 93, 94, 103, 123	Yes	TM	MEM 9		
Rv3705c		CONSERVED HYPOTHETICAL PROTEIN	29, 153, 169, 194, 197, 199, 205	Yes	SP,	CF 2, CF 3, MEM 7, MEM 8, SOL 16		
Rv3706c		CONSERVED HYPOTHETICAL PROLINE RICH PROTEIN	43, 51, 54, 78, 90, 100, 103		TM			
Rv3707c		CONSERVED HYPOTHETICAL PROTEIN	42			MEM 9		
Rv3717		CONSERVED HYPOTHETICAL PROTEIN	15, 24, 48	Yes	SP,	CF 2, CW 14		Mouse 25, Mouse 27
Rv3723		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	35, 97, 108, 109	Yes	TM	MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		Macrophage 23, Mouse 25
Rv3724A	cut5a	PROBABLE CUTINASE PRECURSOR [FIRST PART] CUT5A	33		SP,			
Rv3732		CONSERVED HYPOTHETICAL PROTEIN	34, 90, 117, 122, 127, 149		SP, TM	CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14, WCL 15		
Rv3737		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	228, 233, 238, 292	Yes	TM	MEM 9		
Rv3756c	proZ	POSSIBLE OSMOPROTECTANT (GLYCINE BETAINE/CARNITINE/CHOLINE/L-PROLINE) TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER PROZ	109		TM	MEM 8		
Rv3757c	proW	POSSIBLE OSMOPROTECTANT (GLYCINE BETAINE/CARNITINE/CHOLINE/L-PROLINE) TRANSPORT INTEGRAL MEMBRANE PROTEIN ABC TRANSPORTER PROW	167, 180		TM			Macrophage 24, Mouse 27
Rv3759c	proX	POSSIBLE OSMOPROTECTANT (GLYCINE BETAINE/CARNITINE/CHO	36, 46, 126, 146, 149, 165, 181, 205,	Yes	SP, Lipo	CF 2, CF 3, MEM 7, MEM 9, CW 14		Mouse 27

		LINE/L-PROLINE) BINDING LIPOPROTEIN PROX	248					
Rv3760		POSSIBLE CONSERVED MEMBRANE PROTEIN	44		TM	MEM 7, MEM 9, CW 14		
Rv3763	lpqH	19 KDA LIPOPROTEIN ANTIGEN PRECURSOR LPQH	19, 30, 35, 43, 45, 47, 61, 68, 70, 78, 82, 134	Yes	SP, Lipo	CF 1, CF 2, CF 3, MEM 4, MEM 5, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16	PhoA 18	
Rv3779		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN ALANINE AND LEUCINE RICH	61, 64, 303, 360, 439, 546, 561, 657	Yes	TM	MEM 9, CW 14	BlaTEM 20	
Rv3789		POSSIBLE CONSERVED INTEGRAL MEMBRANE PROTEIN	103		TM	MEM 8		in vitro 29
Rv3792		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	141, 338, 491	Yes	TM	MEM 6, MEM 8, MEM 9, CW 12, CW 14		in vitro 28, in vitro 29
Rv3793	embC	INTEGRAL MEMBRANE INDOLYLACETYLINOSITOL ARABINOSYLTRANSFERASE EMBC (ARABINOSYLINDOLYLACETYLINOSITOL SYNTHASE)	55, 437, 662, 794, 819, 1028		TM	CF 3, MEM 8, MEM 9, MEM 10, CW 13, CW 14		Macaque 22, in vitro 28, in vitro 29
Rv3794	embA	INTEGRAL MEMBRANE INDOLYLACETYLINOSITOL ARABINOSYLTRANSFERASE EMBA (ARABINOSYLINDOLYLACETYLINOSITOL SYNTHASE)	39, 49, 58, 63, 82, 144, 145, 156, 158, 182, 198, 319, 577, 587, 656, 727	Yes	TM	MEM 5, MEM 8, MEM 9, MEM 10, CW 13, CW 14	PhoA 18	Macrophage 23, Mouse 25, in vitro 29
Rv3795	embB	INTEGRAL MEMBRANE INDOLYLACETYLINOSITOL ARABINOSYLTRANSFERASE EMBB (ARABINOSYLINDOLYLACETYLINOSITOL SYNTHASE)	79, 197, 599		TM	MEM 8, MEM 9, CW 14		in vitro 28, in vitro 29
Rv3796	atsH	CONSERVED HYPOTHETICAL PROTEIN	67, 68		Tat SP	CF 2, MEM 8, MEM 9, CW 14		
Rv3802c		PROBABLE CONSERVED MEMBRANE PROTEIN	36, 290, 291		SP, TM	CF 2, CF 3, MEM 5, MEM 8, MEM 9, MEM 10, CW 14	BlaTEM 20	Mouse 27, in vitro 28, in vitro 29
Rv3804c	fbpA, mpt44, 85A	SECRETED ANTIGEN 85-A FBPA (MYCOLYL TRANSFERASE 85A) (FIBRONECTIN-BINDING PROTEIN A) (ANTIGEN 85 COMPLEX A)	39, 76, 127, 129, 160, 288, 330	Yes	SP, Tat SP, TM	CF 1, CF 2, CF 3, MEM 4, MEM 7, MEM 8, MEM 9, MEM 10, CW 12, CW 13, CW 14, WCL 15, SOL 16	PhoA 18	Mouse 27, in vitro 29
Rv3805c		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	417		TM	MEM 9, CW 14		Mouse 25, in vitro 28, in vitro 29
Rv3811	csp	CONSERVED HYPOTHETICAL PROTEIN	15, 38, 178					
Rv3821		PROBABLE CONSERVED INTEGRAL MEMBRANE PROTEIN	62, 74		Tat SP, TM			
Rv3822			72, 74, 90, 152, 185, 249, 325, 341, 358, 362, 368	Yes		CW 14		

Rv3823c	mmpL 8	PROBABLE CONSERVED INTEGRAL MEMBRANE TRANSPORT PROTEIN MMPL8	89, 440		TM	CW 12		Mouse 27, in vitro 29
Rv3831		HYPOTHETICAL PROTEIN	29, 67, 74	Yes	TM			
Rv3835		PROBABLE CONSERVED MEMBRANE PROTEIN	62, 72, 133, 146, 182, 202, 226, 255, 375, 425	Yes	TM	CF 2, CF 3, MEM 6, MEM 7, MEM 8, MEM 9, CW 12, CW 14	BlaTEM 20	
Rv3851		POSSIBLE MEMBRANE PROTEIN	48		SP, TM	MEM 8		
Rv3869	eccB1	POSSIBLE CONSERVED MEMBRANE PROTEIN	70, 88, 130, 463, 476	Yes	TM	CF 2, CF 3, MEM 8, MEM 9, CW 12, CW 14, WCL 15, SOL 16	BlaTEM 20	Mouse 25
Rv3877	eccD1	PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	223, 234		TM	MEM 8, MEM 9, CW 12, CW 14		Mouse 25
Rv3883c	mycP1	MEMBRANE-ANCHORED MYCOSIN MYCP1 (SERINE PROTEASE) (SUBTILISIN-LIKE PROTEASE) (SUBTILASE-LIKE) (MYCOSIN-1)	17, 79, 188, 201, 252, 256, 392	Yes	SP, TM	CF 2, MEM 8, MEM 9, CW 14		
Rv3886c	mycP2	PROBABLE ALANINE AND PROLINE RICH MEMBRANE-ANCHORED MYCOSIN MYCP2 (SERINE PROTEASE) (SUBTILISIN-LIKE PROTEASE) (SUBTILASE-LIKE) (MYCOSIN-2)	170, 510	Yes	SP, TM	CF 3, MEM 5, MEM 9, MEM 10, CW 13, CW 14		
Rv3887c	eccD2	PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	172, 445		TM	MEM 9, CW 14		
Rv3895c	eccB2	PROBABLE CONSERVED MEMBRANE PROTEIN	65, 79, 98, 103, 122	Yes	TM	MEM 8, MEM 9, CW 14		
Rv3901c		POSSIBLE MEMBRANE PROTEIN	54, 67		SP, TM		BlaTEM 20	
Rv3909		CONSERVED HYPOTHETICAL PROTEIN	76, 243, 420	Yes	SP,	MEM 6, MEM 8, MEM 9, CW 14		in vitro 29
Rv3910		PROBABLE CONSERVED TRANSMEMBRANE PROTEIN	70, 353, 372, 430, 1022, 1033, 1049	Yes	Tat SP, TM	CF 3, MEM 4, MEM 5, MEM 6, MEM 8, MEM 9, MEM 10, CW 13, CW 14, SOL 16	BlaTEM 20	Mouse 25, in vitro 29
Rv3912		HYPOTHETICAL ALANINE RICH PROTEIN	177			MEM 9		Macrophage 24

Table 6.1. EXIT Results: all exported proteins and all exported fusion sites. Proteins identified as exported in EXIT in the spleen (Chapter 2) are identified by their genome identification number, name, and function from the NCBI genome annotation (H37Rv RefSeq genome annotation released January 9 2012). Column 4 identifies all statistically enriched fusion sites identified by EXIT as exported in the spleen by the amino acid position of fusion to the 'BlaTEM reporter. Column 5 identifies whether a given protein was also identified as exported in the lungs. Column 6 identifies all *in silico* predicted export signals: signal peptide (SP, (Petersen *et al.*, 2011)), twin-arginine translocation signal peptide (Tat SP, (McDonough *et al.*, 2008)), lipoprotein signal peptide (Lipo, (Sutcliffe & Harrington, 2004)), or transmembrane domain (TM, (Krogh *et al.*, 2001)). Column 7 identifies all previous *in vitro* MS based published methodologies which identified a given protein as exported, with the fractions in which the protein was identified: culture filtrate (CF), membrane (MEM), cell wall (CW), soluble (SOL), or whole cell lysate (WCL). Column 8 identifies all previous genetic reporter based experiments that identified a given protein as exported, with the genetic reporter identified. Column 9 identifies all studies where a

given gene was identified as transcriptionally induced during infection. Numbers given relate to the references and conditions as described below.

Mass spectrometry based approaches to identify exported proteins in fractions: CF: 1 (Rosenkrands *et al.*, 2000), 2 (Malen *et al.*, 2011), 3 (Bell *et al.*, 2012). MEM: 4 (Gu *et al.*, 2003), 5 (Xiong *et al.*, 2005), 6 (Mawuenyega *et al.*, 2005), 7 (Malen *et al.*, 2007), 8 (Malen *et al.*, 2011), 9 (Gunawardena *et al.*, 2013), 10 (Bell *et al.*, 2012). CW: 11 (Mawuenyega *et al.*, 2005), 12 (Wolfe *et al.*, 2010), 13 (Bell *et al.*, 2012). WCL: 15 (Bell *et al.*, 2012). SOL: 16 (Bell *et al.*, 2012). Genetic reporter based approaches to identify exported proteins: 17 (Gomez *et al.*, 2000), 18 (Braunstein *et al.*, 2000), 19 (McDonough *et al.*, 2008), 20 (McCann *et al.*, 2011), 21 (Chapter 2). Genes identified as transcriptionally induced during infection: Macrophage: 30 (Dubnau *et al.*, 2002), 31 (Schnappinger *et al.*, 2003), 32 (Rohde *et al.*, 2007), 33 (BCG) (Rohde *et al.*, 2007), Mouse: 34 (Dubnau *et al.*, 2005), 35 (Talaat *et al.*, 2007), 36 (60 dpi) (Talaat *et al.*, 2007), 37 (45 dpi) (Talaat *et al.*, 2007), Human: 38 (granuloma) (Rachman *et al.*, 2006), 39 (distant lung) (Rachman *et al.*, 2006).

REFERENCES

Bell, C., G.T. Smith, M.J. Sweredoski & S. Hess, (2012) Characterization of the *Mycobacterium tuberculosis* proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. *J Proteome Res* **11**: 119-130.

Braunstein, M., T.I. Griffin, J.I. Kriakov, S.T. Friedman, N.D. Grindley & W.R. Jacobs, Jr., (2000) Identification of genes encoding exported *Mycobacterium tuberculosis* proteins using a Tn552'phoA in vitro transposition system. *J Bacteriol* **182**: 2732-2740.

Dubnau, E., J. Chan, V.P. Mohan & I. Smith, (2005) Responses of *Mycobacterium tuberculosis* to growth in the mouse lung. *Infect Immun* **73**: 3754-3757.

Dubnau, E., P. Fontan, R. Manganelli, S. Soares-Appel & I. Smith, (2002) *Mycobacterium tuberculosis* genes induced during infection of human macrophages. *Infect Immun* **70**: 2787-2795.

Gomez, M., S. Johnson & M.L. Gennaro, (2000) Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. *Infect Immun* **68**: 2323-2327.

Gu, S., J. Chen, K.M. Dobos, E.M. Bradbury, J.T. Belisle & X. Chen, (2003) Comprehensive Proteomic Profiling of the Membrane Constituents of a *Mycobacterium tuberculosis* Strain. *Molecular & cellular proteomics : MCP* **2**: 1284-1296.

Gunawardena, H.P., M.E. Feltcher, J.A. Wrobel, S. Gu, M. Braunstein & X. Chen, (2013) Comparison of the membrane proteome of virulent *Mycobacterium tuberculosis* and the attenuated Mycobacterium bovis BCG vaccine strain by label-free quantitative proteomics. *J Proteome Res* **12**: 5463-5474.

Krogh, A., B. Larsson, G. von Heijne & E.L. Sonnhammer, (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.

Malen, H., F.S. Berven, K.E. Fladmark & H.G. Wiker, (2007) Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics* **7**: 1702-1718.

Malen, H., G.A. De Souza, S. Pathak, T. Softeland & H.G. Wiker, (2011) Comparison of membrane proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra strains. *BMC Microbiol* **11**: 18.

Mawuenyega, K.G., C.V. Forst, K.M. Dobos, J.T. Belisle, J. Chen, E.M. Bradbury, A.R. Bradbury & X. Chen, (2005) *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. *Mol Biol Cell* **16**: 396-404.

McCann, J.R., J.A. McDonough, J.T. Sullivan, M.E. Feltcher & M. Braunstein, (2011) Genome-wide identification of *Mycobacterium tuberculosis* exported proteins with roles in intracellular growth. *J Bacteriol* **193**: 854-861.

McDonough, J.A., J.R. McCann, E.M. Tekippe, J.S. Silverman, N.W. Rigel & M. Braunstein, (2008) Identification of functional Tat signal sequences in *Mycobacterium tuberculosis* proteins. *J Bacteriol* **190**: 6428-6438.

Petersen, T.N., S. Brunak, G. von Heijne & H. Nielsen, (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**: 785-786.

Rachman, H., M. Strong, U. Schaible, J. Schuchhardt, K. Hagens, H. Mollenkopf, D. Eisenberg & S.H. Kaufmann, (2006) *Mycobacterium tuberculosis* gene expression profiling within the context of protein networks. *Microbes Infect* **8**: 747-757.

Rohde, K.H., R.B. Abramovitch & D.G. Russell, (2007) *Mycobacterium tuberculosis* invasion of macrophages: linking bacterial gene expression to environmental cues. *Cell Host Microbe* **2**: 352-364.

Rosenkrands, I., K. Weldingh, S. Jacobsen, C.V. Hansen, W. Florio, I. Gianetri & P. Andersen, (2000) Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. *Electrophoresis* **21**: 935-948.

Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan & G.K. Schoolnik, (2003) Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *J Exp Med* **198**: 693-704.

Sutcliffe, I.C. & D.J. Harrington, (2004) Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiol Rev* **28**: 645-659.

Talaat, A.M., S.K. Ward, C.W. Wu, E. Rondon, C. Tavano, J.P. Bannantine, R. Lyons & S.A. Johnston, (2007) Mycobacterial bacilli are metabolically active during chronic tuberculosis in murine lungs: insights from genome-wide transcriptional profiling. *J Bacteriol* **189**: 4265-4274.

Wolfe, L.M., S.B. Mahaffey, N.A. Kruh & K.M. Dobos, (2010) Proteomic definition of the cell wall of *Mycobacterium tuberculosis*. *J Proteome Res* **9**: 5816-5826.

Xiong, Y., M.J. Chalmers, F.P. Gao, T.A. Cross & A.G. Marshall, (2005) Identification of *Mycobacterium tuberculosis* H37Rv integral membrane proteins by one-dimensional gel electrophoresis and liquid chromatography electrospray ionization tandem mass spectrometry. *J Proteome Res* **4**: 855-861.

APPENDIX III: FUNCTIONAL GENOMICS DATABASE FOR MYCOBACTERIUM TUBERCULOSIS

In an effort to consolidate a vast wealth of data on proteins of *M. tuberculosis* we built a functional genomics database that could become a valuable resource to the *M. tuberculosis* community (https://www.med.unc.edu/microimm/braunsteinlab/research/exported-proteins). Data was compiled for each annotated protein coding gene in the *M. tuberculosis* genome (H37Rv RefSeq genome annotation released January 9 2012) from previous studies identifying exported proteins, *in silico* predicted export signals, essentiality predictions for growth *in vitro* and during infection, and expression data for genes transcriptionally induced during infection.

To aid in predicting the subcellular localization of each protein, data was compiled for previous *in vitro* studies using MS or GR based approaches to identify a given protein as exported, with the fractions in which the protein was identified: culture filtrate (CF), membrane (MEM), cell wall (CW), soluble (SOL), or whole cell lysate (WCL) (Rosenkrands *et al.*, 2000; Malen *et al.*, 2011; Bell *et al.*, 2012; Gu *et al.*, 2003; Xiong *et al.*, 2005; Mawuenyega *et al.*, 2005; Gunawardena *et al.*, 2013; Wolfe *et al.*, 2010; Gomez *et al.*, 2000; Braunstein *et al.*, 2000; McDonough *et al.*, 2008; McCann *et al.*, 2011; Chapter 2). We supplemented this previously published data with data generated by submitting the *M. tuberculosis* proteome to *in silico* prediction programs for export. *In silico* predicted export signals were identified using online prediction programs: signal peptide (Petersen *et al.*, 2011), twin-arginine translocation signal peptide (McDonough *et al.*, 2008), lipoprotein signal peptide (Sutcliffe & Harrington, 2004), or transmembrane domain (Krogh *et al.*, 2001).

To aid in identification of predicted virulence factors results from the multiple high-throughput screens to identify genes essential for *in vitro* growth or virulence in various models of tuberculosis have been included (Dutta *et al.*, 2010; Rengarajan *et al.*, 2005; Stewart *et al.*, 2005; Sassetti & Rubin, 2003; Lamichhane *et al.*, 2005; Zhang *et al.*, 2013; Sassetti *et al.*, 2003). Additionally, results from high-throughput screens to identify genes transcriptionally induced during infection were included (Dubnau *et al.*, 2002; Schnappinger *et al.*, 2003; Rohde *et al.*, 2007; Dubnau *et al.*, 2005; Talaat *et al.*, 2007; Talaat *et al.*, 2007; Rachman *et al.*, 2006).

Because of its usefulness in predicting a function for Rv0199/OmasA (Chapter 3), we analyzed the entire *M. tuberculosis* proteome in Phyre 2.0 (Kelley & Sternberg, 2009) to generate additional functional predictions, and included these hypothetical function predictions alongside annotated functional information from the H37Rv RefSeq genome (released January 9 2012). In this way, we built a database of information for each protein coding gene that can be mined to identify and predict the function of exported proteins and shed light on *M. tuberculosis* pathogenesis. This database can be found at https://www.med.unc.edu/microimm/braunsteinlab/research/exported-proteins.

REFERENCES

Bell, C., G.T. Smith, M.J. Sweredoski & S. Hess, (2012) Characterization of the *Mycobacterium tuberculosis* proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. *J Proteome Res* **11**: 119-130.

Braunstein, M., T.I. Griffin, J.I. Kriakov, S.T. Friedman, N.D. Grindley & W.R. Jacobs, Jr., (2000) Identification of genes encoding exported *Mycobacterium tuberculosis* proteins using a Tn552'phoA in vitro transposition system. *J Bacteriol* **182**: 2732-2740.

Dubnau, E., J. Chan, V.P. Mohan & I. Smith, (2005) responses of *Mycobacterium tuberculosis* to growth in the mouse lung. *Infect Immun* **73**: 3754-3757.

Dubnau, E., P. Fontan, R. Manganelli, S. Soares-Appel & I. Smith, (2002) *Mycobacterium tuberculosis* genes induced during infection of human macrophages. *Infect Immun* **70**: 2787-2795.

Dutta, N.K., S. Mehra, P.J. Didier, C.J. Roy, L.A. Doyle, X. Alvarez, M. Ratterree, N.A. Be, G. Lamichhane, S.K. Jain, M.R. Lacey, A.A. Lackner & D. Kaushal, (2010) Genetic requirements for the survival of tubercle bacilli in primates. *J Infect Dis* **201**: 1743-1752.

Gomez, M., S. Johnson & M.L. Gennaro, (2000) Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. *Infect Immun* **68**: 2323-2327.

Gu, S., J. Chen, K.M. Dobos, E.M. Bradbury, J.T. Belisle & X. Chen, (2003) Comprehensive Proteomic Profiling of the Membrane Constituents of a *Mycobacterium tuberculosis* Strain. *Molecular & cellular proteomics : MCP* **2**: 1284-1296.

Gunawardena, H.P., M.E. Feltcher, J.A. Wrobel, S. Gu, M. Braunstein & X. Chen, (2013) Comparison of the membrane proteome of virulent *Mycobacterium tuberculosis* and the attenuated *Mycobacterium bovis* BCG vaccine strain by label-free quantitative proteomics. *J Proteome Res* **12**: 5463-5474.

Krogh, A., B. Larsson, G. von Heijne & E.L. Sonnhammer, (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.

Lamichhane, G., S. Tyagi & W.R. Bishai, (2005) Designer arrays for defined mutant analysis to detect genes essential for survival of *Mycobacterium tuberculosis* in mouse lungs. *Infect Immun* **73**: 2533-2540.

Malen, H., F.S. Berven, K.E. Fladmark & H.G. Wiker, (2007) Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics* **7**: 1702-1718.

Malen, H., G.A. De Souza, S. Pathak, T. Softeland & H.G. Wiker, (2011) Comparison of membrane proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra strains. *BMC Microbiol* **11**: 18.

Mawuenyega, K.G., C.V. Forst, K.M. Dobos, J.T. Belisle, J. Chen, E.M. Bradbury, A.R. Bradbury & X. Chen, (2005) *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. *Mol Biol Cell* **16**: 396-404.

McCann, J.R., J.A. McDonough, J.T. Sullivan, M.E. Feltcher & M. Braunstein, (2011) Genome-wide identification of *Mycobacterium tuberculosis* exported proteins with roles in intracellular growth. *J Bacteriol* **193**: 854-861.

McDonough, J.A., J.R. McCann, E.M. Tekippe, J.S. Silverman, N.W. Rigel & M. Braunstein, (2008) Identification of functional Tat signal sequences in *Mycobacterium tuberculosis* proteins. *J Bacteriol* **190**: 6428-6438.

Petersen, T.N., S. Brunak, G. von Heijne & H. Nielsen, (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**: 785-786.

Rachman, H., M. Strong, U. Schaible, J. Schuchhardt, K. Hagens, H. Mollenkopf, D. Eisenberg & S.H. Kaufmann, (2006) *Mycobacterium tuberculosis* gene expression profiling within the context of protein networks. *Microbes Infect* **8**: 747-757.

Rengarajan, J., B.R. Bloom & E.J. Rubin, (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci U S A* **102**: 8327-8332.

Rohde, K.H., R.B. Abramovitch & D.G. Russell, (2007) *Mycobacterium tuberculosis* invasion of macrophages: linking bacterial gene expression to environmental cues. *Cell Host Microbe* **2**: 352-364.

Rosenkrands, I., K. Weldingh, S. Jacobsen, C.V. Hansen, W. Florio, I. Gianetri & P. Andersen, (2000) Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. *Electrophoresis* **21**: 935-948.

Sassetti, C.M., D.H. Boyd & E.J. Rubin, (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**: 77-84.

Sassetti, C.M. & E.J. Rubin, (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**: 12989-12994.

Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan & G.K. Schoolnik, (2003) Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *J Exp Med* **198**: 693-704.

Stewart, G.R., J. Patel, B.D. Robertson, A. Rae & D.B. Young, (2005) Mycobacterial mutants with defective control of phagosomal acidification. *PLoS Pathog* **1**: 269-278.

Sutcliffe, I.C. & D.J. Harrington, (2004) Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. *FEMS Microbiol Rev* **28**: 645-659.

Talaat, A.M., S.K. Ward, C.W. Wu, E. Rondon, C. Tavano, J.P. Bannantine, R. Lyons & S.A. Johnston, (2007) Mycobacterial bacilli are metabolically active during chronic tuberculosis in murine lungs: insights from genome-wide transcriptional profiling. *J Bacteriol* **189**: 4265-4274.

Wolfe, L.M., S.B. Mahaffey, N.A. Kruh & K.M. Dobos, (2010) Proteomic definition of the cell wall of *Mycobacterium tuberculosis*. *J Proteome Res* **9**: 5816-5826.

Xiong, Y., M.J. Chalmers, F.P. Gao, T.A. Cross & A.G. Marshall, (2005) Identification of *Mycobacterium tuberculosis* H37Rv integral membrane proteins by one-dimensional gel electrophoresis and liquid chromatography electrospray ionization tandem mass spectrometry. *J Proteome Res* **4**: 855-861.

Zhang, Y.J., M.C. Reddy, T.R. Ioerger, A.C. Rothchild, V. Dartois, B.M. Schuster, A. Trauner, D. Wallis, S. Galaviz, C. Huttenhower, J.C. Sacchettini, S.M. Behar & E.J. Rubin, (2013) Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. *Cell* **155**: 1296-1308.